

**MOBILE ELEMENTE
IM MITOCHONDRIALEN GENOM
DER HEFE:**

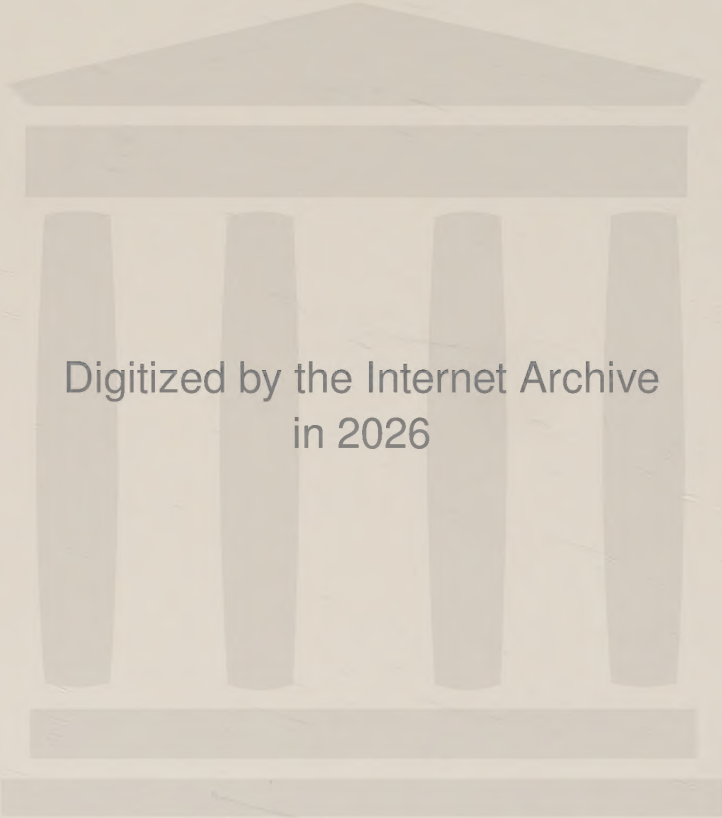
**Hotspotsequenzen der Deletion
und systematische Analyse von Target-
und Insertionssequenzen der GC-Cluster**

Georg F. Weiller

Disbui. München

1987





Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554824_0034

**MOBILE ELEMENTE IM MITOCHONDRIALEN GENOM
DER HEFE:**

**Hotspotsequenzen der Deletion und systematische Analyse
von Target- und Insertionssequenzen der GC-Cluster**

**Inauguraldissertation
zur Erlangung des Doktorgrades der Fakultät Biologie
an der Ludwig-Maximilians Universität München**

**vorgelegt von:
Georg F. Weiller
München im Mai 1987**

1. Berichterstatter: Prof. Dr. R. J. Schweyen

2. Berichterstatter: Prof. Dr. K. Wolf

Tag der mündlichen Prüfung: 22,7.87

INHALT

1.	EINLEITUNG	1
2.	MATERIAL UND METHODEN	10
2.1	MATERIAL	10
2.1.1	Verwendete Materialien	10
2.1.1.1	Enzyme	10
2.1.1.2	Vektoren	10
2.1.1.3	Feinchemikalien	11
2.1.1.4	Radiochemikalien	11
2.1.1.5	Sonstiges	12
2.1.2	Verwendete Stämme	12
2.1.2.1	Sacch.cerevisiae	12
2.1.2.2	E.coli	12
2.2	METHODEN	13
2.2.1	Biochemische Methoden	13
2.2.1.1	Isolation von mtDNA aus Sacch.cerevisiae	13
2.2.1.2	Isolation von Vektor-DNA aus E.coli	13
2.2.1.3	Restriktion	13
2.2.1.4	Gelelektrophorese	14
2.2.1.5	Reisolation von Restriktionsfragmenten	14
2.2.1.6	Nicktranslation	14
2.2.1.7	Radioaktive Markierung von M13-Hybridisierungsproben	14
2.2.1.8	Southern-Analyse	15
2.2.1.9	Koloniehybridisierung	15
2.2.1.10	Radioaktive Endmarkierung von DNA	15
2.2.1.11	Strangtrennung der dsDNA	15
2.2.1.12	Klonierung von mtDNA in E.coli	16
2.2.1.13	DNA-Sequenzanalyse nach der chemischen Methode (Maxam & Gilbert)	16
2.2.1.14	DNA-Sequenzanalyse nach der Kettenabbruch-Methode (Sanger)	16
2.2.1.14.1	Präparation von template-DNA	17
2.2.1.14.2	Hybridisierung von Primer und Template (Annealing)	17

2.2.1.14.3	Kettenverlängerungs-Reaktion	17
2.2.1.14.4	Sequenzgel-Elektrophorese	18
2.2.2	Software-Entwicklung	19
2.2.2.1	Matrix-Programm (MX)	19
2.2.2.1.1	Programmbeschreibung MX	20
2.2.2.1.2	Weitere Programm-Optionen	23
2.2.2.2	Alignment-Programm (ALN)	23
2.2.2.2.1	Programmbeschreibung (ALN)	24
2.2.2.3	Technische Einzelheiten	25
2.2.2.3.1	Software	25
2.2.2.3.2	Hardware	25
3.	ERGEBNISSE UND DISKUSSION	26
3.1	HOTSPOTS DER MITOCHONDRIALEN DELETION	26
3.1.1	Einführung	26
3.1.2	Intron a15b	29
3.1.2.1	Bestimmung der wt-Sequenz im Stamm 777-3A	31
3.1.2.2	GC-Cluster in a15b	35
3.1.3	OXI3-Leader	36
3.1.3.1	Bestimmung der wt-Sequenz im Stamm 777-3A	37
3.1.3.2	GC-Cluster im Bereich des OXI3-Leaders	38
3.1.4	Deletionsklasse A	41
3.1.4.1	Sequenzanalyse	41
3.1.4.2	Mögliche Deletionsmechanismen	43
3.1.5	Deletionsklasse B	45
3.1.5.1	Sequenzanalyse	45
3.1.5.2	Mögliche Deletionsmechanismen	47
3.1.5.2.1	Paarungsmöglichkeit der Deletionsenden	47
3.1.5.2.2	Klasse B-Deletionsenden entsprechen Spleißpunkten	48
3.1.5.2.2.1	Trans-Spleißen	50
3.1.5.2.2.2	Spezifität der Deletionsstelle	52
3.1.5.2.3	Häufige Deletionen an Exon-Intron-Grenzen in unterschiedlichen Organismen	53
3.1.5.2.4	Reverse Transkriptase-Aktivität in Mitochondrien	55
3.1.5.2.5	Mögliche Rolle der RNA bei mitochondrialen Deletionen	56
3.1.5.2.6	Introndeletion ohne die Beteiligung von RNA	60

3.2	GC-CLUSTER	61
3.2.1	Einführung	61
3.2.1.1	Verteilung G+C-reicher Sequenzen in der mtDNA	61
3.2.1.2	Die Targetsite-Hypothese	62
3.2.2	Entwicklung einer GC-Cluster-Systematik	64
3.2.2.1	Methode und Definition	64
3.2.2.2	GC-Cluster Gruppen mit ähnlicher Primärstruktur	65
3.2.2.3	System der GC-Cluster (Ein Vorschlag zur Nomenklatur)	68
3.2.3	Merkmale der GC-Cluster Klassen und Familien	69
3.2.3.1	Die Klasse U	69
3.2.3.1.1	Nukleotidsequenzen der U-Cluster	69
3.2.3.1.2	Targetsequenzen optionaler U-Cluster in A+T-reichen Sequenzen	77
3.2.3.1.3	Organisation der U-Cluster in komplexen Tandemaggregaten	79
3.2.3.1.4	Rekombinationsorte der Cluster-Aggregation	81
3.2.3.1.5	Targetsequenzen optionaler U-Cluster in GC- Clusteraggregaten	82
3.2.3.1.6	Familienbezogene Zusammenfassung aller genannten U- Cluster-Charakteristika	86
3.2.3.1.6.1	Familie U1a	86
3.2.3.1.6.2	Familie U1b	87
3.2.3.1.6.3	Familie U2a	87
3.2.3.1.6.4	Familie U2b	88
3.2.3.1.6.5	Familie U2c	88
3.2.3.1.6.6	Familie U3	89
3.2.3.1.6.7	Familie U4	90
3.2.3.2	Die Klasse G	91
3.2.3.2.1	Nukleotidsequenzen der G-Cluster	91
3.2.3.2.2	Targetsequenzen der G-Cluster-Integration	92
3.2.3.3	Die Klasse V	92
3.2.3.3.1	Nukleotidsequenzen der V-Cluster	92
3.2.3.3.2	Targetsequenzen der V-Cluster-Integration	93
3.2.3.4	Die Klasse I	95
3.2.4	Die Merkmale der drei Hauptklassen im Vergleich	97
3.2.4.1	Exkurs: Vergleich der GC-Cluster-Systematik mit den Ergebnissen von Zamaroczy & Bernardi	99
3.2.5	Mobilität von GC-Clustern	103
3.2.5.1	Hinweise für das mobile Verhalten der Cluster	103
3.2.5.2	Häufigkeit optionaler GC-Cluster	104
3.2.5.3	Insertion versus Deletion	104

3.2.6	Transpositionsmechanismen	105
3.2.6.1	Modelle zur GC-Cluster-Insertion	105
3.2.6.2	Transpositionsmodelle im Vergleich	106
3.2.7	Deletion durch Transposition	110
3.2.7.1	Intramolekulare Transposition	110
3.2.7.2	Mitochondriale Deletionen (mit-delA, rho ⁻)	111
3.3	ZUSAMMENFASSUNG	113
3.3.1	Sequenz-spezifische Deletionen	113
3.3.2	Sequenzierung des wt 777-3A	114
3.3.3	Taget- und Insertiossequenzen der GC-Cluster	115
3.3.4	Insertionsmechanismus der GC-Cluster	117

ABBILDUNGSVERZEICHNISS

Abb.1	Karte des mt Genoms	8
Abb.2	Bildschirmdarstellung einer Homologie-Matrix	21
Abb.3	Bildschirmdarstellung der Häufigkeitsverteilung von Oligonukleotiden	22
Abb.4	Verteilung der OXI3-Deletionsmutanten	27
Abb.5	Intron aI5b-Region: Genomstruktur und Sequenzierungsstrategie	30
Abb.6	Intron aI5b-Region: Nukleotidsequenz	32
Abb.7	OXI3-Leaderregion: Genomstruktur und Sequenzierungsstrategie	37
Abb.8	OXI3-Leaderregion: Nukleotidsequenz	39
Abb.9	mit-delA: Genomstruktur und Sequenzierungsstrategie	42
Abb.10	mit-delA-Nahtstelle: Sequenz im Vergleich mit dem wt	43
Abb.11	mit-delB: Genomstruktur und Sequenzierungsstrategie	45
Abb.12	mit-delB-Nahtstelle: Sequenz im Vergleich mit dem wt.	47
Abb.13	Mögliche Intrastrangpaarung der Klasse B-Deletionsorte	48
Abb.14	Sequenzhomologie der rechten Deletionsstelle in aI5b mit der 5'Spleißsequenz des Introns aI1	49
Abb.15	Klasse B-Deletion durch Fehlspleißen	49
Abb.16	Struktur der mit-delB-RNA nach einem hypothetischen Fehlspleißen	57
Abb.17	Heteroduplex zwischen mtDNA und fehlgespleißter RNA	59
Abb.18	GC-Cluster Polymorphismen im OXI3-Bereich	63
Abb.19	Homologiematrix der GC-Cluster und ihrer Konsenssequenzen	66
Abb.20	U1-Cluster	70
Abb.21	U2-Cluster	72
Abb.22	U3-Cluster	74
Abb.23	U4-Cluster	76
Abb.24	Targetsequenz optionaler U-Cluster in A+T reichen Bereichen	78
Abb.25	GC-Clusteraggregate	80
Abb.26	Mögliche Rekombinationsorte der GC-Cluster-Aggregation	81

Abb.27	GC-Cluster-Polymorphismen im COB-Intron 3 und nach dem 15S rRNA-Gen: Genomstruktur und Sequenzierungsstrategie	83
Abb.28	Targetsequenzen optionaler U-Cluster in GC-Clusteraggregaten	85
Abb.29	G-Cluster	91
Abb.30	V-Cluster	93
Abb.31	Targetsequenzen der V-Cluster	94
Abb.32	I-Cluster	96
Abb.33	Dinukleotid-Verdoppelung der drei GC-Cluster-Klassen	97
Abb.34	Charakteristika der drei GC-Cluster-Klassen im Vergleich	98
Abb.35	Vergleich der GC-Cluster-Systematik mit einem von Zamaroczy & Bernardi parallel entwickelten Einteilungssystem	101
Abb.36	Modelle zu GC-Cluster-Integration	105
Abb.37	Symmetrische replikative Transposition	108
Abb.37	Intramolekulare Transposition	110

ABKÜRZUNGEN UND GENETISCHE BEZEICHNUNGEN

A	Adenosin
ATP	Adenosintriphosphat
bp	Basenpaar (-e)
C	Cytosin
CAMV	Cauliflower Mosaikvirus
cDNA	revers transkribiert mRNA
CRF	unterbrochener Leserahmen (closed reading frame)
CTP	Cytosintriphosphat
dATP	3'desoxi-ATP
dCTP	3'desoxi-CTP
ddNTP	2',3'didesoxi-NTP
DNA	Desoxiribonukleinsäure
ds	Doppelstrang
E.coli	Escherichia coli
EDTA	Ethylendiamintetraacetat
G	Guanosin
GC ⁺	Bezeichnung für ein Allel, welches ein gegebenes GC-Cluster enthält
GC ⁻	Bezeichnung für ein Allel, welches ein gegebenes GC-Cluster nicht enthält
GTP	Guanosintriphosphat
hnRNA	ungespleißte Ribonukleinsäure im Zellkern (heterogenous nuclear RNA)
IPTG	Isopropylthiogalaktosid
IS	Insertions-Element
ITP	Inosintriphosphat
kb, kbp	Kilobasen, -paare
LB	flüssiges Nährmedium mit Trypton, Hefeextrakt, NaCl und Glycerin
mit-	Punktmutation oder kurze Deletion in der mitochondrialen DNA

mit-delA	mitochondriale Deletionsmutante der Klasse A
mit-delB	mitochondriale Deletionsmutante der Klasse B
mRNA	Boten-Ribonukleinsäure (messenger RNA)
mt	mitochondrial (-e, -en)
N	A, T, C oder G (Nukleotid)
<i>N.crassa</i>	<i>Neurospora crassa</i>
NTP	Nukleotid-Triphosphat
ORF	offener Leserahmen (open reading frame)
<i>P.anserina</i>	<i>Podospira anserina</i>
PEG	Polyethylenglykol
rho-	Makrodeletion der mtDNA, wobei der verbleibende Genomrest bis zur vollen Genomlänge reiteriert wird.
rho ^o	Mutante, die das gesamte mt Genom verloren hat
rpm	Umdrehungen pro Minute (revolutions per minute)
RNA	Ribonukleinsäure
<i>Sacch.cerevisiae</i>	<i>Saccharomyces cerivisiae</i>
<i>Schizosach.pombe</i>	<i>Schizosaccharomyces pombe</i>
SDS	Natriumlaurylsulfat (sodium dodecyl sulfate)
ss	Einzelstrang (single strand)
T	Thymidin
TBE	Tris-Borat-EDTA (Puffer)
TE	Tris-EDTA (Puffer)
TEMED	N,N,N',N',-Tetramethylethylendiamin
Tris	Tris(hydroxymethyl)-aminomethan
tRNA	Transferribonukleinsäure
U	Unit (Einheit der Enzymaktivität)
UE	Untereinheit (von Enzymkomplexen)
wt	Wildtyp
ÜN	Über Nacht
X-Gal	5-Brom-4-Clor-3Indolyl-βD-Galaktosid

1. EINLEITUNG

"Die grundlegende biologische Invariante ist die DNA. Mendels Definition des Gens als des invarianten Trägers der Vererbungsmerkmale, die chemische Identifikation des Gens durch Avery, die durch Hershey bestätigt wurde, und die Aufhellung der strukturellen Grundlagen seiner replikativen Invarianz durch Watson und Crick stellen deshalb ohne jeden Zweifel die wichtigsten Entdeckungen dar, die jemals in der Biologie gemacht wurden. Dem ist noch die Evolutionstheorie anzufügen, die übrigens nur durch diese Entdeckungen zu ihrer vollen Bedeutung und Bestätigung gelangen konnte."

Aus Jacques Monod: «Le hasard et la nécessité»
Paris 1970

Zufällige Mutationen der weitgehend starren DNA stellen in Kombination mit der Selektion den Motor der Evolution dar. Ort und Häufigkeit der Mutationen entscheiden über die Entwicklungsfähigkeit genetischer Systeme. Die Entwicklung von Elementen innerhalb der DNA, welche auf das Mutationsgeschehen selbst Einfluß nehmen können, stellt daher einen wichtigen Multiplikator der Evolutionsgeschwindigkeit dar.

Die Geschwindigkeit, mit der die Evolution voranschreitet, läßt sich auf zweierlei Weise abschätzen. Eine indirekte Methode besteht darin, die Anzahl der unterschiedlichen Aminosäuren des selben Proteins in unterschiedlichen Arten mit der Zeit zu vergleichen, welche seit der Aufspaltung beider Arten vergangen ist. Eine direkte Methode besteht in der Beobachtung von Veränderungen in wachsenden Zellen (Bakterien-, Drosophila-, Mauskolonien etc.). Die Anzahl der pro Generation aufgetretenen Mutationen kann daraus berechnet werden.

Beide Methoden führen zu dem gleichen Ergebnis: zur Veränderung eines bestimmten Nukleotids sind etwa 10^9 Generationen nötig. Diese durchschnittliche Mutationsgeschwindigkeit hat sich in dem überschaubaren Zeitabschnitt (mehrere Millionen Jahre) nicht verändert.

Eine grobe Abschätzung am Beispiel von Drosophila soll zeigen, daß die Evolution rezenter Organismen nicht durch unabhängige Veränderungen einzelner

Nukleotide möglich war. Die Generationszeit der Taufliege beträgt etwa 14 Tage, ihre Keimbahn kann damit innerhalb eines Jahres 25mal repliziert werden. Selbst dann, wenn man die nötigen Umwege der Evolution nicht berücksichtigt, würde die Anpassung der 10^8 Nukleotide des Drosophilagenoms durch unabhängige Einzelmutationen wenigstens 4×10^{15} Jahre in Anspruch nehmen. Zu lange für eine Erde, die erst seit etwa 5×10^9 Jahren besteht.

Die Aufklärung von Protein- und DNA-Strukturen ermöglichte es die Strategie der Evolution im Wesentlichen zu erkennen. So ist heute sicher, daß sich viele Gene aus bereits vorhandenen entwickelt haben. Viele Beispiele verwandter Proteine wurden gefunden. Die drei Serinproteasen Chymotrypsin, Trypsin und Elastase erfüllen unterschiedliche Funktionen im Stoffwechsel, dennoch sind 40% der Aminosäuren in den drei Proteinen unverändert. 54 der 124 Aminosäuren des α -Lactoglobins der Milch sind identisch mit denen des Enzyms Lysozym im Hühnereiß. Die Sequenzen einer Vielzahl unterschiedlicher Gene lassen deren Abstammung von gemeinsamen Vorfahren zweifelsfrei erkennen.

Die komplizierten Reaktionen, welche Enzyme als Biokatalysatoren bewerkstelligen, verlangen meist viele einzelne Reaktionsschritte. Im typischen Fall wird jeder dieser Schritte von einem abgeschlossenen Bereich des Enzyms (Domäne) katalysiert. So besitzt die Phosphoglycerat-Kinase, deren Tertiärstruktur durch Röntgenstrukturanalyse aufgeklärt wurde, zwei Domänen. Während eine der Domänen die Bindung von ATP bewirkt, ist die zweite für die Bindung von 3-Phosphoglycerat zuständig. Nach der Bindung beider Substrate erfolgt eine Konformationsänderung, wobei die beiden Domänen zusammentreten und die aktive Stelle der Katalyse bilden.

Die universellen Cosubstrate (ATP, NADH, etc.) werden von einer großen Zahl unterschiedlicher Enzyme gebunden. Durch die Kombination verschiedener, bereits funktionierender Domänen konnten viele neue Gene entwickelt werden.

Als weitere Beispiele, welche die Zusammenfassung unabhängiger Domänen zu neuen Proteinen belegen, seien Immunglobuline, Albumine und Collagene genannt. In diesen Fällen werden die einzelnen Domänen von eigenständigen Genabschnitten (Exons) codiert, welche von intervenierenden Sequenzen (Introns) unterbrochen werden. Es ist wahrscheinlich, daß diese Introns die Neukombination der Exons erleichtert haben, da ohne die Introns die Kombination der Exons

nur innerhalb eines sehr kurzen DNA-Bereiches zu sinnvollen neuen Proteinen geführt hätte.

Die Evolution neuer Gene vollzieht sich in der Regel durch die Diversifizierung eines bereits vorhandenen Gens oder durch die Kombination mehrerer vorhandener Genabschnitte (Domänen): häufig durch beides. Die Weiterentwicklung einer Nukleotidsequenz wird dadurch erleichtert, daß viele Gene in mehrfacher Kopienzahl im Genom vorhanden sind. Die zusätzlichen Kopien (Pseudogene) sind für den Organismus nicht notwendig, sie bieten der Evolution damit ein weites Experimentierfeld. Neue noch nicht optimierte Gene werden so nicht sofort zum Opfer der Selektion. Genverdoppelung ist ein wichtiges Prinzip der Evolution.

Die DNA der meisten Eukaryonten enthält viele oft wiederholte Sequenzen. Hochrepetitive DNA (Satelliten-DNA) besteht aus sehr kurzen (z.B. 5 bp) Sequenzmotiven welche meist 10^6 bis 10^7 mal wiederholt werden. In vielen Säugern bildet ein einziger Typ dieser DNA bereits mehr als 10% des gesamten Genoms. Die Satelliten DNA unterliegt einer raschen Änderung im Laufe der Evolution. Das menschliche Genom enthält mindestens drei verschiedene Satelliten-DNAs. Unterschiedliche Mixturen werden an allen Centromeren gefunden. Vergleicht man beispielsweise die Sequenzen zweier Homologer (väterliches und mütterliches Chromosom), so weichen die Satelliten-DNAs normalerweise völlig voneinander ab. Es ist nicht bekannt, ob dieser DNA eine Aufgabe in der Zelle zukommt, sie könnte möglicherweise die Stabilität eines Chromosoms erhöhen. Da sie vorwiegend um den Bereich des Centromers der Chromosomen vorkommt, ermöglichen Rekombinationen dieser DNAs, ganze Chromosomenarme neu zu kombinieren.

Mittelrepetitive Sequenzen unterscheiden sich von den o.g. hochrepetitiven Sequenzen in verschiedener Hinsicht:

(i) Sie sind etwa 130 bis 150 bp lang und in hundert bis millionenfacher Kopienzahl im Genom verstreut. (ii) Zumindest einige der mittelrepetitiven Sequenzen werden transkribiert. (iii) Lokalisation und Sequenz dieser DNAs verändern sich im Laufe der Evolution langsamer als bei Satelliten-DNAs, jedoch schneller als bei normalen Genen.

Eukaryonten besitzen viele unterschiedliche Familien von mittelrepetitiver DNA. Die im Menschen häufigste Familie (Alu-Familie) besteht aus einer Sequenz von

ca. 300 bp, welche im Genom etwa 300.000 mal auftritt. Auch mittelrepetitiver DNA kann heute keine notwendige Funktion zugeordnet werden. Da sie funktionierende Gene von einander trennt und miteinander rekombinieren kann, ermöglicht sie die Veränderung der Chromosomenorganisation. Die Nachbarschaft bestimmter Gene und ihre Lage auf dem Chromosom ist eine wichtige Voraussetzung für die Koordination der Gene im Organismus.

Es wird heute angenommen, daß ein Großteil aller repetitiven Sequenzen keine unmittelbare Funktion für die Zelle besitzen, sondern sich in parasitärer Weise auf ihrem Wirtsgenom ausbreiten (Egoistische DNA).

DNA-Elemente, deren einzige Funktion in ihrer eigenen Verbreitung zu bestehen scheint, sind in nahezu allen Genomen weit verbreitet. Zu diesen Elementen zählen neben den o.g. repetitiven Sequenzen Viren, transponierbare Elemente und Introns. Sie sind in der Regel mehr oder weniger mobil, d.h. sie können ihren Aufenthaltsort im Genom wechseln. Manche dieser Elemente codieren Proteine, welche ihre Verbreitung unterstützen. Hohe Mobilität fördert die Ausbreitung der Sequenzen, die damit einhergehenden Veränderungen im Genom können jedoch den Tod der Zelle bewirken. Dieses Problem wird von den verschiedenen Elementen unterschiedlich gelöst.

Repetitive Sequenzen sind darauf angewiesen, sich auf nicht-codierende Sequenzen zu beschränken. Introns haben dagegen die Fähigkeit entwickelt, sich nach der Transkription aus der RNA zu entfernen (Spleißen), wodurch es ihnen möglich wird, codierende Sequenzen zu unterbrechen. Eine Vielzahl transponierbarer Elemente in Prokaryonten und Eukaryonten codieren Proteine, welche ihre Transposition regulieren. Dadurch kann der Schaden begrenzt werden, den zu häufige Transposition im Wirtsgenom bewirken würde. Manche Elemente inserieren bevorzugt in spezifische Targetsequenzen. Hierdurch können sie die Wahrscheinlichkeit verringern, in Sequenzen zu inserieren, welche für das Fortbestehen der Zelle nötig sind. Viele können sich auch außerhalb des Wirtsgenoms als eigenständige zirkuläre DNA (Plasmide) replizieren. Viren sind darüberhinaus noch in der Lage, Proteinhüllen auszubilden, welche ihre DNA beim Verlassen der Zelle schützt und die Infektion neuer Zellen erleichtert. Da Vermehrung und Freisetzung der Viren häufig zum Tod der Zelle führt, haben viele von ihnen (temperente Viren) die Fähigkeit entwickelt, nach ihrer Integration ins Wirtsgenom ihre eigene

Vermehrung zu unterdrücken. Erst wenn Streßfaktoren den möglichen Tod der Zelle ankündigen, erwacht die Aktivität des Virus.

Die Grenzen zwischen all diesen Elementen sind nur schwer zu ziehen. So kann der Bakteriophage (Bakterienvirus) Mu auch als transponierbares Element auftreten. Die Ty-Elemente der Hefe und copia-Elemente von *Drosophila* stellen Zwischenformen zwischen Transposons und Retroviren (RNA-Viren) dar (39). Viroide (zirkuläre pathogene Pflanzen-RNAs) und Virusoide (Satelliten-RNA in Pflanzen) sind mit Introns verwandt (29). Introns können als Plasmide auftreten (83) oder als Transposons, wobei sie die zur Transposition nötigen Proteine (Transposasen) selbst codieren (58, 68)). Die hochrepetitiven Sequenzen des Schleimpilzes *Physarum polycephalum* verhalten sich wie eukaryontische Transposons (91).

Das parasitäre Verhalten vieler dieser egoistischen DNA-Elemente ist im Laufe der Evolution einem symbiontischen Verhältnis gewichen.

In Prokaryonten sind zwei benachbarte IS-Elemente (Insertionselemente) in der Lage, nicht nur sich selbst zu replizieren, sondern auch das dazwischenliegende Gen. Der gesamte Komplex aus Wirts-DNA und den beiden IS-Elementen ist dann als autonom replizierendes zirkuläres Plasmid in der Lage die Wirtsinformation zu exprimieren. Die Fusion mehrerer Plasmide ermöglicht es verschiedene Gene auf einem Plasmid zusammenzufassen. Als bekannte Beispiele sind die Resistenz-Transfer-Faktoren zu nennen. Hierbei handelt es sich um Plasmide, auf welchen die Information für mehrere Antibiotikaresistenzen zusammengefaßt ist. Sie ermöglichen den Bakterien nicht nur die zur Resistenz nötigen Proteine in großen Mengen zu synthetisieren, sondern sind auch in der Lage, die Zellen zu wechseln und andere nicht-resistente Bakterienstämme zu immunisieren. Dadurch wird eine enorme Steigerung der Evolutionsgeschwindigkeit erreicht.

Gene können somit auf selektive Weise von einem Organismus auf einen zweiten übertragen werden, ohne daß ein Generationswechsel benötigt wird (horizontale Vererbung). Damit kann auch die Grenze überwunden werden, die der vertikalen Vererbung durch die Spezies gesetzt ist.

Viele der mobilen DNA Sequenzen stellen unverzichtbare Elemente des Organismus dar, wodurch sich eine klare Grenze zwischen egoistischer DNA und anderen Genen nicht ziehen läßt. Hier ist z.B. der Wechsel des Paarungstyps der

Hefe *Sacch.cerevisiae* zu nennen. Haploide Hefezellen können zwei verschiedene Geschlechtsformen (a und α) besitzen. Nur unterschiedliche Paarungstypen sind in der Lage, sich zu Zygoten zu vereinen. Bei einigen Hefestämmen ist der Paarungstyp stabil, bei anderen wechselt er fast in jeder Zellgeneration. Ermöglicht wird dieser rasche Wechsel dadurch, daß beide Informationen (a und α) in den Zellen vorhanden sind. Eine definierte Stelle in der DNA (Paarungstyp-Locus) entscheidet über den tatsächlichen Paarungstyp. Sowohl das a -Gen, als auch das α -Gen stellen springende Elemente dar, welche in den Paarungstyp-Locus inserieren können und die dort vorhandene Information ersetzen. Für diesen als 'Kassettenmechanismus' bezeichneten Transpositionsmechanismus gibt es weitere Beispiele.

Trypanosomen, die Erreger der Schlafkrankheit, sind einzellige Parasiten des Menschen und anderer Säuger. Sie werden im Blut des Menschen von Antikörpern attackiert. Die Immunabwehr, welche sich gegen die Oberflächenproteine des Parasiten wendet, benötigt einige Tage zu ihrer Formierung. Durch einen raschen Wechsel ihrer Oberflächenproteine sind Trypanosomen in der Lage, dem Immunsystem lange Zeit zu entgehen. Ähnlich dem Paarungstyp-Locus der Hefe entscheidet hier ein spezieller Transkriptionslocus, welches Gen exprimiert wird. Trypanosomen halten die Gene für etwa 100 verschiedene Oberflächenproteine bereit. Sie alle stellen mobile Elemente dar, welche an den Expressionslocus transponiert werden können.

Manche transponierbaren Elemente (Inversionselemente) sind in der Lage ihren Lokus zu verlassen und in entgegengesetzter Orientierung wieder zu integrieren. Einige von ihnen tragen Transkriptionsstart-Sequenzen (Promotoren). Ihre Orientierung entscheidet, ob die Gene links oder rechts des Inversionselements exprimiert werden. Das Bakterium *Salmonella typhimurium* nützt solch ein System (hin), um eines von zwei verschiedenen Oberflächenproteinen zu bilden und damit die antigenen Eigenschaften seiner Oberfläche (H-Antigene) zu verändern. Die Bakteriophagen Mu und P1 regulieren durch ähnliche Systeme (gin bzw. cin) ihre Wirtsspezifität. Weitere Beispiele könnten hinzugefügt werden (als Übersichtsartikel vgl. (105)).

Verschiedene mobile Elemente sind darauf spezialisiert, miteinander zu rekombinieren, was zum spezifischen Verlust des dazwischenliegenden DNA-Bereiches und zur Verknüpfung (Ligation) der Randsequenzen führt (sequenz-spezifische

Deletion). Durch die Ligation verschiedener Domänen können in rascher Folge neue Proteine erzeugt werden. Eine weitere Möglichkeit verschiedene Proteindomänen zu verknüpfen, besteht in der Fähigkeit mancher Introns, nicht nur sich selbst, sondern auch weitere, von diesen Introns flankierte Exons aus der transkribierten RNA zu spleißen (differentielles Spleißen). Das Immunsystem von Säugern nutzt beide Mechanismen (sequenz-spezifische Deletion und differentielles Spleißen), um die Vielzahl der Antikörper zu erzeugen, welche für die Immunabwehr benötigt werden.

Sequenz und Organisation mobiler Elemente lassen auf die Verwandtschaft vieler dieser Elemente schließen. Eine besondere Bedeutung scheint dabei den Rändern der Elemente zuzukommen. Diese werden häufig durch zu einander palindrome Sequenzen gebildet. Dadurch können die Randsequenzen miteinander in Paarung treten. Die dazwischenliegenden Sequenzen bilden so eine haarnadelförmige Schleife (Hairpinstruktur). Durch diese Struktur werden sowohl die Ränder des Elements als auch die Ränder der Wirts-DNA in räumliche Nähe gebracht. Die Annäherung der betroffenen Sequenzen bildet eine Voraussetzung zur sequenz-spezifischen Rekombination.

Sequenzen, an denen Deletionen oder Insertionen besonders häufig auftreten (Hotspots), wurden in unterschiedlichen genetischen Systemen bestimmt (1, 10, 32, 37, 47, 51, 52, 61, 101, 116, 119). Besonderheiten an diesen Sequenzen ermutigten manche Autoren (1, 10, 26, 32, 47, 51, 52, 110) erste Modelle des molekularen Geschehens zu entwerfen. Illegitime sequenz-spezifische Rekombination, Replikation und reverse Transkription prozessierter RNA erscheinen heute als zelluläre Vorgänge, welche Insertionen und Deletionen längerer Nukleotidsequenzen bewirken.

Als besonders geeignetes System, das Insertions- und Deletionsgeschehen zu beobachten, ist das mitochondriale Genom von *Sacch.cerevisiae* anzusehen. Das Genom beinhaltet ausschließlich Informationen, die zur Synthese einiger weniger Komponenten der Atmungskette benötigt werden. Unter geeigneten Kulturbedingungen kann *Sacch.cerevisiae* auf dieses Genom vollständig verzichten (ρ^0 Stämme). Auf fermentierbarem Substrat sind weder mit-Mutationen (Punktmutationen und kleine Deletionen in der mtDNA) noch ρ^- (Deletion von mehr als 50% des mt Genoms) noch ρ^0 -Mutationen letal. Beliebige Mutanten können damit weiter kultiviert und daher analysiert werden.

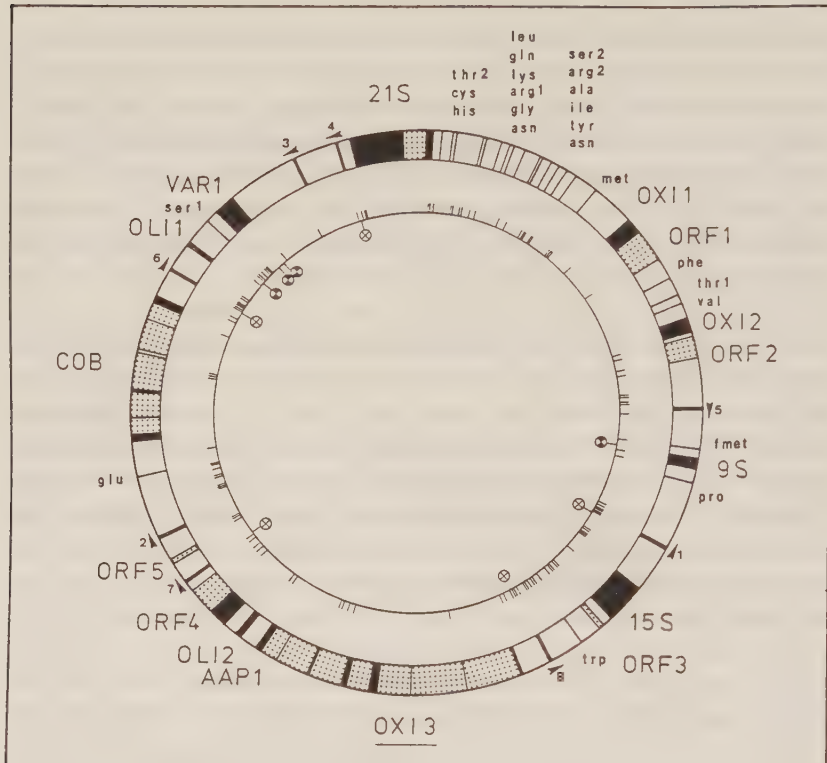


Abb.1 Karte des mt Genoms

Exons sind durch dunkle, Introns und intergenische ORFs durch gepunktete Balken wiedergegeben. Radiale Linien geben die Positionen der tRNAs wieder. Die Replikationsstart Sequenzen (ORI1 bis 8) sind mit Pfeilen markiert, welche zugleich ihre relative Orientierung im Genom wiedergeben. Die Gene OXI1, 2 und 3 codieren die UE II, III, und I der Cytochrom-Oxidase; COB, Cytochrom b; AAP1, OLI2 und OLI1, UE 8, 6 und 9 der ATPase; VAR1 codiert ein Protein welches mit der kleinen UE der mt Ribosomen assoziiert ist. 15S und 21S sind die Gene für die kleine und große rRNA. 9S codiert eine RNA welche an der Prozessierung der tRNAs beteiligt ist. GC-Cluster werden in dieser Arbeit in die Klassen U, G und V eingeteilt. Der innere Kreis gibt die Positionen von GC-Clustern wieder. U-Cluster weisen nach außen; G-Cluster sind mit ⊗, V-Cluster mit ⊕ markiert.

Die mtDNA zählt zu den instabilsten Genomen, die bekannt sind. Sie enthält eine Reihe optionaler Elemente, wie kurze G+C-reiche Sequenzen (GC-Cluster), A+T-reiche Abschnitte und fakultative Introns. Alle diese optionalen Sequenzen unterliegen keinem merkbaren Selektionsdruck. In Abb. 1 ist die Karte eines mitochondrialen Genoms dargestellt, welches alle optionalen Elemente enthält.

Die Tatsache, daß einige Mutagen bevorzugt mit mtDNA interagieren, führt in Verbindung mit der relativ geringen Größe des mt Genoms (75-85kb), sowie besonderen in Mitochondrien zusätzlich möglichen Kartierungstechniken (rho-Retensionskartierung) (98-100, 106) dazu, daß die mt Genkarte durch eine große Zahl von gut charakterisierten Mutanten dicht abgedeckt ist. Ein überraschend hoher Anteil (30%) aller mit-Mutanten kartiert dabei im OXI3-Gen (18, 99, 106, Sang Ho Kim pers. Mittlg.). Feinkartierung von mit-Mutanten im OXI3-Genbereich (18, Sang Ho Kim, pers. Mittlg.) zeigt, daß dies u.a. auf zwei in dieser Region besonders häufig auftretende Deletionstypen zurückzuführen ist. Die beiden Mutationen wurden mit 'mit-delA' (Klasse A) und 'mit-delB' (Klasse B) bezeichnet (18). Beide Deletionsklassen stellen je etwa 7% aller mit-Mutationen des Genoms.

Ein Teil der vorliegenden Arbeit betrifft die Aufklärung der Primärstruktur dieser beiden Hotspots der Deletion. Er versucht zum Verständnis des Mechanismus beizutragen, welcher solchen Deletionen zu Grunde liegt. Die Deletionen der Klasse A treten an GC-Clustern auf, welche ihrerseits optionale Elemente darstellen. Diese im mt Genom weit verbreiteten repetitiven Sequenzen (vgl. Abb.1) gaben wiederholt Anlaß zu der Vermutung, es könne sich hierbei um mobile Elemente handeln. Dieser Frage wurde in einem weiteren Teil dieser Arbeit nachgegangen. Zunächst wurde eine Systematik der GC-Cluster entwickelt, welche es erlaubt, die Cluster detailliert zu beschreiben und zu gruppieren. Die Bestimmung von Sequenzen, welche sich nur durch die Anwesenheit von GC-Clustern unterscheiden, erlaubte es, eine Hypothese (Targetsite-Hypothese) zu entwickeln, welche die minimalen Voraussetzungen einer Sequenz beschreibt, als Zielsequenz (Akzeptor) der GC-Cluster-Transposition zu fungieren. Die Aufklärung der Verwandtschaftsbeziehungen unter den GC-Clustern, und der systematische Vergleich mit ihren möglichen Insertionssequenzen ermöglichten es, einen wahrscheinlichen Transpositionsmechanismus der GC-Cluster vorzuschlagen. Er weist Analogien zu jenem anderer mobiler Elemente auf und könnte eine der möglichen Ursachen für die bemerkenswerte Instabilität des mt Genoms darstellen.

2. MATERIAL UND METHODEN

2.1 MATERIAL

2.1.1 VERWENDETE MATERIALIEN

2.1.1.1 ENZYME

DNA-Ligase	Boehringer, Mannheim
DNA-Polymerase I	BRL, Maryland (USA)
Klenow-Polymerase	Boehringer, Mannheim
Phosphatase, alkalische	Boehringer, Mannheim
Polynukleotidkinase	Boehringer, Mannheim
Proteinase K	Boehringer, Mannheim
Restriktionsenzyme	BRL, Maryland (USA)
	Boehringer, Mannheim
	Biolaps, Beverly (USA)
S1-Nuklease	Boehringer, Mannheim
Zymolase 5000	Kirin Brewery Co. Ltd. (Japan)

2.1.1.2 VEKTOREN

pUC8	pUC9		
M13mp8	M13mp9 (75)	M13mp18	M13mp19 (74)

2.1.1.3 FEINCHEMIKALIEN

Soweit nicht anders vermerkt, wurden alle Chemikalien von der Fa. Merck, Darmstadt bezogen.

Acrylamid	Serva, Heidelberg
Agar	Difco, Detroit (USA)
Agarose	Seakem, Maine (USA)
Ammoniumpersulfat	Serva, Heidelberg
Ampicillin	Serva, Heidelberg
Bisacrylamid	Serva, Heidelberg
Bisbenzimid	Riedel de Haen, Hanover
Caesiumchlorid	Baker, Denver (Holland)
DMS	Fulka, Buchs (Schweiz)
DNA aus Heringssperma	Böehringer, Mannheim
EDTA	Sigma, St.Louis (USA)
Ethidiumbromid	Serva, Heidelberg
Formamid	Fulka, Buchs (Schweiz)
Glucose	Hedinger, Stuttgart
Hefeextrakt	Difco, Detroit (USA)
Hydrazin	Eastman Organic Chemicals
Hybridisierungsprimer	Böehringer, Mannheim
IPTG	Böehringer, Mannheim
Piperidin	Sigma, St.Lois (USA)
Polyethylenglycol	Serva, Heidelberg
Sequenzierungsprimer	Böehringer, Mannheim
SDS	Serva, Heidelberg
Spermidin	Sigma, St.Lois (USA)
TEMED	Serva, Heidelberg
Tris	Sigma, St.Lois (USA)
tRNA aus E.coli	Böehringer, Mannheim
X-Gal	Böehringer, Mannheim

2.1.1.4 RADIOCHEMIKALIEN

$\alpha^{32}\text{P}$ -dNTP, $\gamma^{32}\text{P}$ -dNTP	Amersham-Buchler, Braunschweig
$\alpha^{35}\text{S}$ -dATP	New England Nuclear, Dreieich

2.1.1.5 SONSTIGES

Agfapan 100 Film	Agfa-Gevaert, Leverkusen
3MM-Papier	Schleicher und Schüll, Dassel
Nitrozellulose	Schleicher und Schüll, Dassel
Röntgenfilm	Kodak und Dupont
Röntgen-Verstärkerfolie	CAWO, Schrobenuhausen

2.1.2 VERWENDETE STÄMME

2.1.2.1 SACCH.CEREVISIAE

777-3A (62)	511-4A	2393A
D8-25A	DBY746	DBY747
DL1	FL100	FM9A
IC166-7A	IC125-10C	IC8
KL14-4A (100)	M12/G4	PB2-1C
RM590-9D	SM11-6C	

2.1.2.2 E.COLI

JM101	Δ <u>lacpro</u> , <u>supE</u> , <u>thi</u> ; F' <u>traD36</u> , <u>proAB</u> , <u>lacIq</u> , Z Δ M15
BMH71-18	Δ <u>lacpro</u> , <u>supE</u> , <u>thi</u> ; F' <u>lacIq</u> , <u>proAB</u> , Z Δ M15

2.2 METHODEN

2.2.1 BIOCHEMISCHE METHODEN

2.2.1.1 ISOLATION VON mTDNA AUS SACCH.CEREVISIAE

Große Mengen mtDNA wurden aus Mitochondrien extrahiert und durch isopyknische Zentrifugation in CsCl gereinigt (64).

Zur Gewinnung kleiner Mengen mtDNA wurden 1,5 ml einer ÜK abzentrifugiert (12000 rpm) in Sorbitolpuffer (1,5M) resuspendiert und mit Zymolase 60000 protoplastiert (20 Min., 37°C). Die Protoplasten wurden in Tris-HCl (50 mM, pH 7,8), EDTA (50 mM), SDS (1%) in Anwesenheit von ProteinaseK (50 µg/ml) lysiert. Nach wiederholter Phenolextraktion wurde die DNA mit Ethanol aus dem Lysat gefällt.

2.2.1.2 ISOLATION VON VEKTOR-DNA AUS E.COLI

Große Mengen DNA von pBR322, pUC und M13(RF) wurden durch alkalische Lyse (70) gewonnen und wie dort angegeben gereinigt.

Zur Gewinnung kleiner Mengen von Plasmid-DNA wurde die Methode von Birnboim & Doly (12) angewandt.

2.2.1.3 RESTRIKTION

Bei der Restriktion von DNA wurden die Empfehlungen der Hersteller bezüglich der einzelnen Restriktionsenzyme beachtet.

2.2.1.4 GELELEKTROPHORESE

Die Auftrennung der Restriktionsfragmente erfolgte meist in horizontalen 0,6-2%igen Agarosegelen (5-8 V/cm, TBE-Puffer). Zur Auftrennung kleiner Restriktionsfragmente (<300 bp) wurden 12-30 cm lange vertikale 4-8%ige Acrylamid-Gele (10-20 V/cm, TBE-Puffer) verwendet.

Die Gele wurden mit Ethidiumbromid (1 µg/ml) angefärbt und im UV-Durchlicht (254 nm) photographiert (Agfapan 100).

2.2.1.5 REISOLATION VON RESTRIKTIONSFRAGMENTEN

Die aus Restriktionsgelen im UV-Licht ausgeschnittenen Banden wurden teilweise in Dialyseschläuche eingeschlossen, teilweise in eine 'Biotrap'-Elutionskammer gelegt (Schleicher & Schüll). Die Extraktion der DNA erfolgte durch Elektroelution (10-15 V/cm, TBE-Puffer).

2.2.1.6 NICKTRANSLATION

Die radioaktive Markierung der DNA erfolgte nach der Methode von Rigby et al. (96), jedoch ohne Zugabe von DNase. Nicht inkorporierte ^{32}P -Triphosphate wurden über eine Sephadex-G50-Säule abgetrennt.

2.2.1.7 RADIOAKTIVE MARKIERUNG VON M13-HYBRIDISIERUNGSPROBEN

Zur Herstellung von Hybridisierungsproben aus bereits in M13 klonierten DNA Fragmenten wurde ein spezieller Hybridisierungs-Primer an das 3'Ende des Polylinkers hybridisiert und in Anwesenheit von $\alpha^{32}\text{P}$ -dATP mit Klenow-Polymerase verlängert (54)

2.2.1.8 SOUTHERN-ANALYSE

Die Southern-Analyse von mtDNA wurde entsprechend der von Southern et al. entwickelten Methode durchgeführt (111). Zur Hybridisierung der Nitrozellulose-Filter wurden weitere Methoden verwendet (54, 60).

2.2.1.9 KOLONIEHYBRIDISIERUNG

Wenn anstelle eines definierten Restriktionsfragment ein Gemisch vieler Fragmente in E.coli kloniert wurde, so wurde zur Identifikation des gewünschten Klons ein Abklatsch des Koloniemusters mit geeigneten DNA-Proben hybridisiert. Diese Koloniehybridisierung wurde entsprechend der Methode von Grundstein & Hogness (48) durchgeführt.

2.2.1.10 RADIOAKTIVE ENDMARKIERUNG VON DNA

Die zurückversetzten 3'Enden von dsDNA wurden mit dem Klenow-Fragment der E.coli DNA-Polymerase und $\alpha^{32}\text{P}$ -dNTP aufgefüllt (70).

Zur Markierung glatt geschnittener oder vorstehender 5'Enden wurden das 5'Phosphat mit alkalischer Phosphatase entfernt. Zur Phosphorylierung der entstandenen 5'Hydroxyl-Gruppe wurde anschließend T4-Polynukleotidkinase und $\alpha^{32}\text{P}$ -ATP benutzt (70).

2.2.1.11 STRANGTRENNUNG DER dsDNA

An beiden Enden markierte dsDNA-Fragmente wurden vor ihrer Sequenzierung nach der Methode von Maxam & Gilbert (72) in ihre Einzelstränge aufgetrennt.

2.2.1.12 Klonierung von mtDNA in E.coli

Vor der Ligation wurden aus Agarose isolierte Restriktionsfragmente im 'Elutip' (Schleicher und Schüll) gereinigt. Am 5'Ende überstehende Nukleotide (MboII-Verdau) wurden zunächst mit S1-Nuklease abgedaut und anschließend mit Klenow-Polymerase aufgefüllt (70). M13 bzw. pUC Vektoren wurde am Polylinker mit den passenden Restriktionsenzymen gespalten. Im Falle glatter Restriktionsschnittstellen wurden die Vektoren zusätzlich mit Kalbsthymus-Phosphatase dephosphoryliert (70). Die Ligation von mtDNA mit der Vektor-DNA erfolgte nach der von Maniatis et al. (70) empfohlenen Methode.

Die E.coli Stämme JM101 bzw. BMH71-18 wurden mit dem Ligationsgemisch transformiert. Die Transformation von M13-Vektoren erfolgte nach der Methode von Messing et al. (75), bei der Transformation von pUC-Vektoren wurde die Methode von Mandel & Higa (69) angewandt.

2.2.1.13 DNA-Sequenzanalyse nach der chemischen Methode (Maxam & Gilbert)

Endmarkierte DNA-Fragmente wurden den chemischen Spaltreaktionen, wie von Maxam & Gilbert (72, 73) beschrieben, unterzogen und in 50 cm langen denaturierenden Gelen (4-20% Acrylamid) aufgetrennt. Die Gele wurden einem Kodak XR 5-Röntgenfilm zwischen vier Tagen und fünf Wochen exponiert.

2.2.1.14 DNA-Sequenzanalyse nach der Kettenabbruch-Methode (Sanger)

Die DNA-Sequenzierung mit Hilfe von Substratanalogons (ddNTP) welche die Kettenverlängerung während der DNA-Synthese basen-spezifisch inhibieren, wurde 1977 von F. Sanger et al. (97) entwickelt. Die Einführung von genetisch manipulierten M13-Phagen (74, 75) macht es möglich, auch lange DNA-Sequenzen in verhältnismäßig kurzer Zeit zu bestimmen.

2.2.1.14.1 PRÄPARATION VON TEMPLATE-DNA

Zur Gewinnung einzelsträngiger Phagen-DNA wurden 3ml LB-Medium mit M13 infizierten E.coli Zellen (weißer Plaque) angeimpft und ÜN bei 37°C unter heftigem Schütteln inkubiert. Anschließend wurden die Bakterienzellen abzentrifugiert (12000 rpm) und die Phagen mit PEG 6000 aus dem Überstand gefällt. Nach Phenolextraktion der in TE resuspendierten Phagen konnte ihre ssDNA mit Ethanol gefällt werden (Ausbeute ca. 5 µg DNA/10 ml Medium).

2.2.1.14.2 HYBRIDISIERUNG VON PRIMER UND TEMPLATE (ANNEALING)

In 5 µl TF wurden etwa 1 µg Template-DNA gemeinsam mit der äquimolaren Menge Primer-DNA (ca. 2,5 ng) auf 55°C erhitzt und langsam (2h) auf 30°C im Wasserbad abgekühlt.

2.2.1.14.3 KETTENVERLÄNGERUNGS-REAKTION

Je ein Viertel des Annealing-Gemisches wurde mit einem dNTP/ddNTP Gemisch, 0,3 U Klenow-Polymerase und $\alpha^{32}\text{P}$ -dATP bzw. ^{35}S -dATP vermengt und 15-20 Min. bei 30°C inkubiert. Anschließend wurden alle vier dNTPs im Überschuß zugesetzt und die Reaktion nach 15 Min. mit Formamid gestoppt. Die Sequenzierungen in Gegenwart von $\alpha^{32}\text{P}$ -dATP erfolgten nach dem Protokoll von Messing et al. (75). Bei den Reaktionen mit ^{35}S -dATP wurde auf ein Protokoll der Firma NEN zurückgegriffen.

Bei der Sequenzierung von GC-Clustern (Klasse U2a) konnten die o.g. Methoden nicht immer zufriedenstellen, da diese G+C-reichen Palindrome wegen ihrer starken Rückfaltungsneigung zur Kompression der Banden während des Gellaufes führen. Daher wurden dGTP und ddGTP durch die Substratanaloga dITP und ddITP ersetzt (42). Im Gegensatz zu Guanosin kann Inosin nur zwei Wasserstoffbrücken mit Cytosin ausbilden.

2.2.1.14.4 SEQUENZGEL-ELEKTROPHORESE

Die Reaktionsgemische wurden kurz denaturiert (2 Min., 100°C) und mit ca. 3000V auf 50 cm langen denaturierenden (50% Harnstoff) Polyacrylamid-Gelen elektrophoretisch aufgetrennt. Die Gele hatten Polyacrylamid-Konzentrationen zwischen 6% und 16%, ihr Durchmesser variierte zwischen 0,2 mm und 0,5 mm. Zur Minimierung von Bandenkompressionen wurden beheizbare Glasplatten benützt, und die Temperatur während des Gellaufes bei 60°C gehalten (3). Um das Auflösungsvermögen der Gele zu erhöhen, waren die Gelplatten so präpariert, daß das Gel an eine der Platten kovalent gebunden wurde und nach dem Lauf getrocknet werden konnte (3).

2.2.2 SOFTWARE-ENTWICKLUNG

Die mtDNA von *Sacch.cer.* beherbergt weit über 100 G+C-reiche Cluster, von denen viele identische oder ähnliche Sequenzen besitzen. Sie werden in Abschn. 3.2 ausführlich behandelt. Zu Beginn dieser Arbeit konnte bereits angenommen werden, daß diese GC-Cluster aus nur wenigen typischen Sequenzmotiven zusammengesetzt sind. Ein Teil dieser Arbeit bestand darin, diese Grundmotive zu erfassen und damit die verwandtschaftlichen Beziehungen unterschiedlicher GC-Cluster zu untersuchen. Die Auswertungen wurden mit Hilfe eines Microcomputers durchgeführt.

Für Microcomputer besteht heute ein vielseitiges Angebot an Software, welche in der Lage ist, Homologiebeziehungen verschiedener Sequenzen zu untersuchen. Die zur Analyse der GC-Cluster benötigte Software muß in der Lage sein, etwa 200 kurze Sequenzen zu verwalten und untereinander zu vergleichen. Sie muß kurze, in ähnlicher Form häufig wiederholte Sequenzmotive aufdecken und darüberhinaus genügend Flexibilität besitzen, die Bedeutung charakteristischer Sequenzpositionen zu berücksichtigen. Sie sollte erlauben, die kleinstmögliche Anzahl von Konsenssequenzen abzuleiten, welche nötig ist, das Gros der GC-Cluster zu beschreiben. Zur sukzessiven Optimierung dieser Konsenssequenzen muß deren Homologie mit den einzelnen GC-Clustern schnell und übersichtlich erfaßt werden können.

Da bisher für Microcomputer kein Programm zur Verfügung stand, welches allen o.g. Bedingungen entsprach, wurde die nötige Software selbst entwickelt (Programme MX und ALN).

2.2.2.1 MATRIX-PROGRAMM (MX)

Das Programm sollte in der Lage sein, homologe Teilbereiche zu erkennen, einzeln abweichende Basen, ebenso wie kleinere Insertionen und Deletionen zu berücksichtigen und die Vielzahl der Homologiebeziehungen möglichst vollständig und dennoch überschaubar darzustellen. Hierfür bot sich die graphische Darstellung in Form einer Punktmatrix an. Ein entsprechender Algorithmus wurde von Tinokio et al. (117) erstmals vorgestellt.

Eine Vielzahl von Punktmatrix-Programmen wurde bisher entwickelt. Sie alle stellen die zu vergleichenden Sequenzen als Abszisse und Ordinate einer Matrix dar. Wird beim Vergleich von horizontaler und vertikaler Sequenz ein frei wählbares Homologiekriterium (z.B. mindestens 8 identische Nukleotide von 10 Nukleotiden) erreicht, so führt dies zu einem Punkt in der Matrix. Die Koordinaten des Punktes geben dabei die Positionen der Homologen in der vertikalen und horizontalen Sequenz wieder. Längere homologe Bereiche erscheinen damit als diagonale Linien. Insertionen bzw. Deletionen innerhalb des homologen Bereiches führen lediglich zur parallelen Verschiebung der weiterführenden diagonalen Linie. Da dieses Verfahren für Insertionen bzw. Deletionen besonders sensitiv ist, wird es häufig zum Vergleich längerer Protein- und Nukleotidsequenzen herangezogen.

2.2.2.1.1 PROGRAMMBESCHREIBUNG MX

Bei der Anwendung des Punktmatrix-Algorithmus zur detaillierten Analyse von mehr als 100 kurzen GC-Cluster Sequenzen ergaben sich folgende Probleme:

(i) Die Darstellung einer homologen Sequenz durch einen einzelnen Punkt gibt nicht die exakten Endpunkte des homologen Bereiches wieder. Der Algorithmus wurde daher so modifiziert, daß sämtliche identischen Basen innerhalb eines homologen Bereiches in einem Matrixpunkt resultieren, einzelne Fehlpaarungen jedoch zu einer Unterbrechung der Diagonalen führen. Die Interpretation der Matrix wird so erheblich vereinfacht.

(ii) Der Vergleich von 140 GC-Clustern würde zu 140^2 (19600) einzelnen Graphiken führen. Diese zu vergleichen würde den Anwender überfordern. Das Programm muß daher sämtliche Einzelmatrizen in einer übergeordneten Matrix zusammenfassen. Ein Matrixplot, bei dem 137 GC-Cluster mit den bereits ausgearbeiteten Konsenssequenzen verglichen werden, ist in Abschn. 3.2.2.2 (Abb.19) wiedergegeben.

Zur Entwicklung dieser Konsenssequenzen war es nötig, die konkreten Sequenzen eines häufig wiederkehrenden Homologen zu erfassen. Prinzipiell kann die konkrete Sequenz eines Homologen auf Grund der Koordinaten seiner Diagonalen bestimmt werden. Da das Zurückverfolgen vieler Homologen ein zeitaufwendiges

Unterfangen ist, wurde in das Programm eine Option implementiert, welche das Auffinden der Sequenz unterstützt. Dazu kann die Punktmatrix auch auf dem Bildschirm dargestellt werden, wobei der Bildschirm ein bewegliches Fenster in der virtuellen Gesamtmatrix darstellt (vgl. Abb.2).

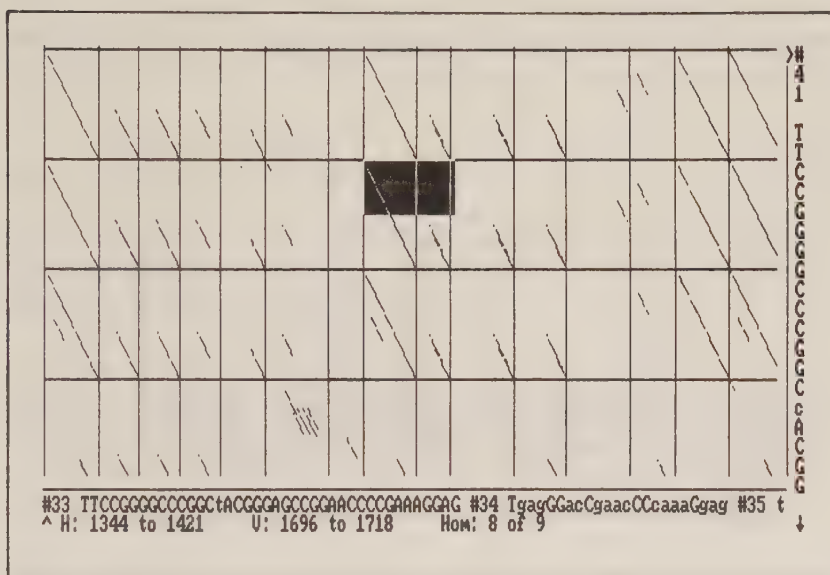


Abb.2 Bildschirmdarstellung einer Homologie-Matrix

(Siehe Text)

Innerhalb des Bildschirmfensters befindet sich ein kleineres, bewegliches Fenster (Abb.2 Mitte) mit dessen Hilfe beliebige Diagonalen (Homologen) angewählt werden können. Die konkrete Sequenz des entsprechenden Homologen wird dann am Bildschirmrand dargestellt. Übereinstimmende Nukleotide werden in Großbuchstaben wiedergegeben, die Nummer der Cluster sowie weitere Parameter werden ebenfalls angezeigt.

Um die Signifikanz homologer Teilbereiche zu beurteilen, ist es nötig zu erfassen, wie häufig ein beliebiges Oligonukleotid in einer Vielzahl von Sequenzen (Clustern) vorhanden ist. Dies wird am einfachsten erreicht durch die Addition aller Matrix-Punkte, die in den Spalten über der Sequenz des Oligonukleotids stehen. Das Programm ist in der Lage, diese Addition auszuführen und das Ergebnis in Form einer Graphik in den Bildschirm einzublenden.

In Abb.(3) ist die Darstellung einer Häufigkeitsverteilung von Oligonukleotiden wiedergegeben

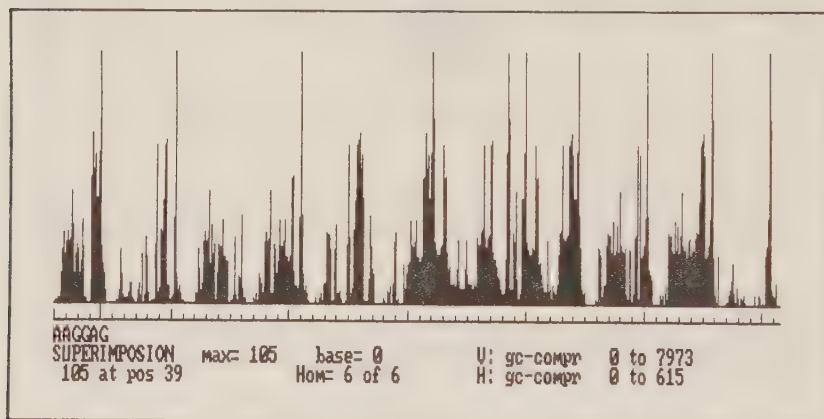


Abb.3 Bildschirmdarstellung der Häufigkeitsverteilung von Oligonukleotiden

Der Bildschirm stellt folgende Informationen bereit: Die maximale Häufigkeit eines eindeutigen Oligonukleotids mit 6 bp Länge (Hom= 6 of 6 bp) beträgt 105 (max= 105). Die Häufigkeiten aller Oligonukleotide der horizontalen Sequenz (Filename 'gc-compr') zwischen den Positionen 0 und 615 (0 to 615) wurde in einem längeren Abschnitt der identischen vertikalen Sequenz gesucht. Der Zeiger in der Linealzeile steht auf Position 39. Das sich dort befindende Oligonukleotid AAGGAG wurde 105 mal gefunden (105 at pos 39) (siehe Text).

Die Ordinate gibt die Häufigkeit des Oligonukleotids wieder, seine Position wird durch die Koordinaten der Abszisse festgelegt. Die konkreten Sequenzen der Oli-

gonukleotide werden auf der darunter liegenden Bildschirmzeilen wiedergegeben. Sie können durch einen entlang der Abszisse beweglichen Zeiger selektiert werden.

Das MX-Programm ermöglichte es, auf verhältnismäßig einfache und rasche Weise häufig wiederkehrende Sequenzmotive in GC-Clustern zu erkennen und sie den einzelnen Clustern zuzuordnen. Daraus konnten die Konsenssequenzen der unterschiedlichen Clustertypen abgeleitet, und die Cluster in distinkte Klassen und Familien eingeteilt werden.

2.2.2.1.2 WEITERE PROGRAMM-OPTIONEN

Weitergehende Informationen zu MX werden im Anschluß an diese Arbeit publiziert (Nucl. Acids Res. in Vorbereitung). Im folgenden werden nur einige allgemeine Optionen des MX-Programms angesprochen.

Horizontale und vertikale Sequenzfiles sind frei wählbar, sie können mit jedem Textverarbeitungsprogramm ediert werden. Damit können beliebige Nukleotid- bzw. Proteinsequenzen verglichen werden. Da die vertikale Sequenz auch in ihre komplementäre Sequenz überführt werden kann, ist das Programm auch geeignet 'Hairpinstrukturen' zu untersuchen.

Der Abbildungsmaßstab des Matrix-Plots ist in vier Stufen veränderbar. Auf einer DIN A4-Seite können max. 1950 x 1950 bp dargestellt werden. Größere Bereiche werden automatisch auf mehrere überlappende Papierbahnen aufgeteilt.

Innerhalb der zu untersuchenden Sequenzen können beliebige Markierungen angebracht werden. Die resultierende Punktmatrix wird an diesen Stellen von Gitterlinien überlagert.

2.2.2.2 ALIGNMENT-PROGRAMM (ALN)

Um homologe Sequenzen sinnvoll untereinander zu reihen (Alignment) müssen meist kleinere Insertionen und Deletionen eingeführt werden, wobei die Information des deletierten Bereiches nicht verloren gehen darf. Homologe Nukleotide

sollten sich in der Darstellung von nicht-homologen Nukleotiden unterscheiden (Groß-Kleinschrift etc.). Das Anfertigen umfangreicher Sequenztabelle gestaltet sich mit normalen Textverarbeitungsprogrammen sehr schwierig, da die dazu wichtigen Funktionen von diesen Programmen nicht unterstützt werden. Zur Erleichterung solcher Arbeiten und zur Optimierung der mit Hilfe von MX gefundenen Konsenssequenzen entwickelte ich das Programm ALN.

2.2.2.2.1 PROGRAMMBESCHREIBUNG (ALN)

Das Programm verwaltet eine unbegrenzte Anzahl kurzer Sequenzen (max. 255 bp) in einer Datei. Jede dieser Sequenzen kann in einem zweiten Feld (Kommentarfeld) näher beschrieben werden (Nummer, Genort, Typ, etc.). Nach Angabe einer beliebigen Kombination von Ein- und Ausschlußkriterien (z.B. die An- und Abwesenheit bestimmter Oligonukleotide oder Vermerke im Kommentarfeld) werden die damit selektierten Sequenzen auf dem Bildschirm dargestellt. Weitere können hinzugefügt werden.

Die Sequenzen lassen sich entsprechend einer wählbaren und veränderbaren Konsenssequenz ordnen, wobei homologe Bereiche stets hell (invers) dargestellt sind. Die dazu nötigen Insertionen und Deletionen können von Hand oder automatisch durchgeführt werden. Bereiche, die durch Deletion verloren gingen, werden mitverwaltet. Eine kurze Beschreibung (Kommentarfeld) der aktuell bearbeiteten Sequenz erscheint in einem separaten Bildschirmfenster. Weiterhin kann das Programm Sequenzen einzeln invertieren (Umwandlung in die Sequenz des komplementären DNA-Stranges) und beliebige Sequenzen nach dem Grad ihrer Homologie sortieren. Die zur weiteren Bearbeitung nötigen Funktionen wie Blockbewegen, -deletieren, -abspeichern, -einfügen etc. sind vorhanden.

Die Sequenztabelle kann auf einem Drucker ausgegeben werden. Nicht homologe Basen erscheinen dabei in Kleinbuchstaben, deletierte Bereiche sind vermerkt. Breitere Tabellen werden auf mehrere Papierbahnen aufgeteilt. Die fertigen Sequenztabelle können auch in einem Standardformat (ASCII) abgespeichert und somit mit herkömmlichen Textverarbeitungsprogrammen weiterbearbeitet werden. Die Abb.20-23, Abb.25 und Abb.30 dieser Arbeit wurden so erzeugt.

2.2.2.3 TECHNISCHE EINZELHEITEN

2.2.2.3.1 SOFTWARE

Zur leichten Anpassungsfähigkeit an spezielle Aufgaben wurden die Programme MX und ALN größtenteils in GWBASIC® (Microsoft Basic) geschrieben und nicht kompiliert. Um dennoch eine maximale Ausführungsgeschwindigkeit zu erreichen, wurden alle zeitaufwendigen Vergleichs- und Darstellungsroutinen in Assembler (Maschinensprache) programmiert. Das MX-Programm ist damit z.B. in der Lage, bei einem Homologiegrad bei dem 10 von 11 bp identisch sind in nur 45 Sekunden 1000 x 1000 bp zu vergleichen. Der Ausdruck des Matrix-Plots wird nur durch die Geschwindigkeit des Druckers begrenzt. Die max. Länge der mit MX untersuchbaren Sequenz beträgt 60 000 bp. ALN benötigt für den Homologievergleich aller auf dem Bildschirm dargestellten Sequenzen mit einer Konsenssequenz (inklusive der inversen Darstellung der homologen Nukleotide) nur ca. 1/10 Sekunde.

Beide Programme werden dem 'public domain'-Sektor zu den üblichen Bedingungen unentgeltlich zur Verfügung gestellt.

2.2.2.3.2 HARDWARE

Die Programme laufen auf allen MS-DOS Rechnern (IBM® PC/AT oder kompatibel) mit mindestens 256 Kb RAM. Die momentanen Versionen (MX V1.1, ALN V2.0) unterstützen die Color-Graphik Karte. Die Unterstützung der höher auflösenden Herkules®-Graphik-Karte ist in Vorbereitung. Der Ausdruck ist mit allen IBM®- oder Epson®-kompatiblen Druckern möglich.

3. ERGEBNISSE UND DISKUSSION

3.1 HOTSPOTS DER MITOCHONDRIALEN DELETION

3.1.1 EINFÜHRUNG

Das OXI3-Gen, in dessen Bereich alle hier behandelten mitochondrialen Deletionsmutanten (mit-del) kartieren, ist das längste Gen in den Mitochondrien der Bäckerhefe *Sacch.cerevisiae*, es codiert die UE I der Cytochrom-Oxidase. Dieses Mosaikgen wird von bis zu neun Introns unterbrochen, wobei die Introns aI5a und aI5b nur in den sog. 'langen' Genomen auftreten.

Verschiedene Untersuchungen an mitochondrialen Mutanten erbrachten, daß 50-60% aller mit- im OXI3-Bereich kartieren (18, 106, 98-100, Sang Ho Kim pers. Mittlg.), obwohl diese Genregion nur ca. 10% des mt Genoms einnimmt.

Unsere Kollektion mitochondrialer Mutanten enthält derzeit etwa 200 gut charakterisierte Mutanten mit Defekten in der OXI3-Region. Alle sind durch Mn^{2+} Mutagenese im wt 777-3A induziert und in einer Weise isoliert worden, daß sie als Ergebnis unabhängiger Mutationen gelten können (R. J. Schweyen, pers. Mittlg.). Unter diesen Mutanten befinden sich 60 (=30%) mit langen Deletionen (18, Sang Ho Kim, pers. Mittlg.). Dies ist insofern erstaunlich, als der Anteil solcher Deletionen in anderen mt Genen unter 5% bleibt (3% in OXI1 (124), 1,5% in COB (49, R. J. Schweyen, pers. Mittlg.). Es war daher zu vermuten, daß besondere Strukturen innerhalb des OXI3-Gens das Auftreten langer Deletionen begünstigen. Genetische Feinkartierung konnte diese Vermutung bestätigen. Die Deletionsenden der OXI3-Mutanten verteilten sich auf nur wenige 'Hotspots'. Die Deletionsmutanten wurden daher einzelnen Deletionsklassen mit einer jeweils unterschiedlichen Zahl identischer Mitglieder zugeordnet (18). Diese Kartierungen wurden später erweitert und in vielen Fällen durch Restriktionsanalyse bestätigt (Sang Ho Kim, pers. Mittlg.).

Abb.4 gibt die Verteilung der 60 mit-del Mutanten und das Ausmaß ihrer Deletionen wieder.

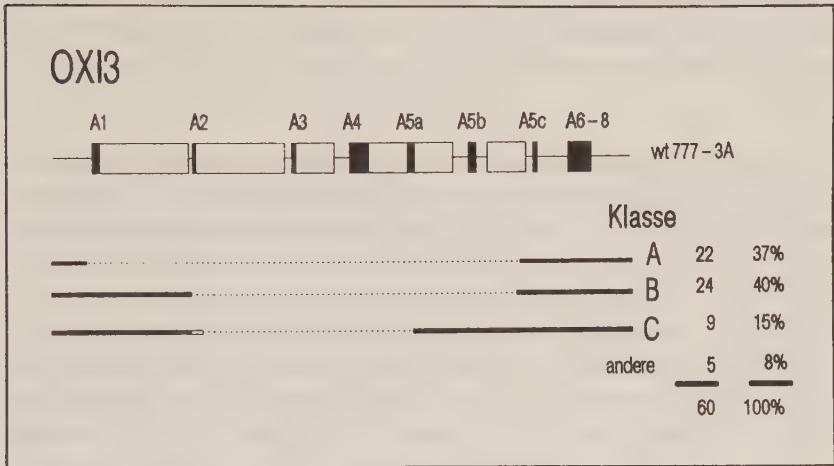


Abb.4 Verteilung der OXI3-Deletionsmutanten

Exons und Introns des OXI3-Gens sind durch dunkle bzw. helle Balken dargestellt. Die Unterbrechungen in den dicken Linien darunter geben die in den Deletionsklassen A, B und C deletierten Genbereiche wieder. Etwa 50-60% aller mit-Mutanten betreffen das OXI3 Gen, 30% davon sind lange Deletionen. Die Verteilung der 60 OXI3-Deletionsmutanten unserer Mutantenkollektion auf die einzelnen Klassen ist in absoluten und relativen Zahlen angegeben.

Die beiden Deletionsklassen A und B stellen je etwa 6% aller mit-Mutationen. Ihr häufiges Auftreten weist auf einen sequenz-spezifischen Deletionsmechanismus hin.

Ein Ziel dieser Arbeit war es, die molekulare Architektur dieser beiden dominierenden Hotspots (Klasse A und B) zu analysieren. Dies konnte durch die Bestimmung der Nahtstellensequenzen von sieben mit-del-Mutanten und deren Vergleich mit wt-Sequenzen erreicht werden.

Die OXI3-Sequenz des kurzen wt-Stammes D273-10B wurde von Bonitz et al. (16) bestimmt, die Sequenzdaten der beiden fakultativen Introns aI5a und aI5b des langen Genoms von KL14-4A wurden von Hensgens et al. (50) nachgereicht. Die Leadersequenz des Gens lag bisher nicht vollständig vor. Während das 3' Ende des Leaders in den Sequenzdaten von D273-10B enthalten ist, konnte der 5' Teil in KL14-4A sequenziert werden (89). Dazwischen klaffte eine Sequenzlücke von ca. 150 bp.

Erste Sequenzierungen von Klasse A-Mutanten ergaben, daß ihr linkes Deletionsende in diese Sequenzlücke fällt, während ihr rechtes Deletionsende ein optionales G+C-reiches Element (GC-Cluster) betrifft, welches in der vorliegenden wt-Sequenz (KL14-4A) nicht vorhanden ist. Damit wurde es nötig die Nukleotidsequenz an den betroffenen Bereichen im wt 777-3A, der Stammform der mit-del-Mutanten, zu bestimmen. Die für die Bereiche um den OXI3-Leader und das Intron aI5b erstellten Restriktionskarten ergaben, daß sich die Schnittstellen der beiden Restriktionsenzyme HpaII (CCGG) und HaeIII (GGCC) in den beiden wt-Stämmen KL14-4A und 777-3A erheblich unterscheiden. Bezüglich anderer (nicht in GC-Clustern schneidender) Restriktionsenzyme ergaben sich kaum Unterschiede zwischen den beiden wt-Stämmen. Es war daher wahrscheinlich, daß 777-3A mehrere optionale GC-Cluster enthält, wodurch die Ausdehnung der Sequenzierungsarbeiten auf die entsprechenden Bereiche lohnend erschien. Auf Grund der Restriktionsdaten beider Stämme sind weitere GC-Cluster-Dimorphismen innerhalb des OXI3-Gens unwahrscheinlich.

3.1.2 INTRON aI5b

Das fakultative Intron aI5b ist ebenso wie die Intron aI5a, bI1, bI2 und bI3 nur in sog. 'langen', mtDNAs der Hefe vorhanden. Das zur Gruppe II gehörende Intron aI5b besitzt einen langen offenen Leserahmen (ORF), dessen Sequenzhomologie mit anderen mt ORFs die Codierung einer 'RNA-Maturase' vermuten läßt. Der Leserahmen ist in zweifacher Hinsicht unüblich:

(i) Der ORF schließt nicht an den Leserahmen des vorausgehenden Exons an, wie dies bei allen übrigen intronischen ORFs proteincodierender mt Gene der Fall ist; er gleicht damit den freistehenden ORFs (ORF1 bis 5) bzw. dem intronischen ORF des 21S rRNA Gens (ω -Intron) (109).

(ii) Der ORF enthält kein einziges ATG-Codon. Die Translation des Leserahmens könnte entweder durch einen zusätzlichen Spleißschritt, welcher den ORF an das vorangehende Exon ligiert, erreicht werden, oder bei einem anderen Codon beginnen (50). Die Existenz unüblicher Startcodons konnte in anderen Organismen nachgewiesen werden (2, 11, 36, 125).

Die Restriktionsdaten des wt 777-3A unterscheiden sich in diesem Genbereich von jenen des bereits sequenzierten wt KL14-4A in vielfacher Hinsicht. Neben einer fehlenden BglII-Schnittstelle besitzt 777-3A mehrere zusätzliche HpaII-HaeIII-Schnittstellen, weshalb die Anwesenheit optionaler GC-Cluster vermutet wurde (vgl. Abb.5). Die 3'Deletionsendpunkte der beiden aus dem wt 777-3A abgeleiteten mit-Deletionsklassen A und B fallen in diesen ORF, wobei ein mit-delA-Deletionsendpunkt eines dieser optionalen Cluster betrifft. Sowohl zur Bestimmung der Deletionsstelle von Klasse A-Mutanten, als auch zum Studium der exakten Insertionsstelle optionaler GC-Cluster, wurde daher der entsprechende Genbereich im wt 777-3A sequenziert.

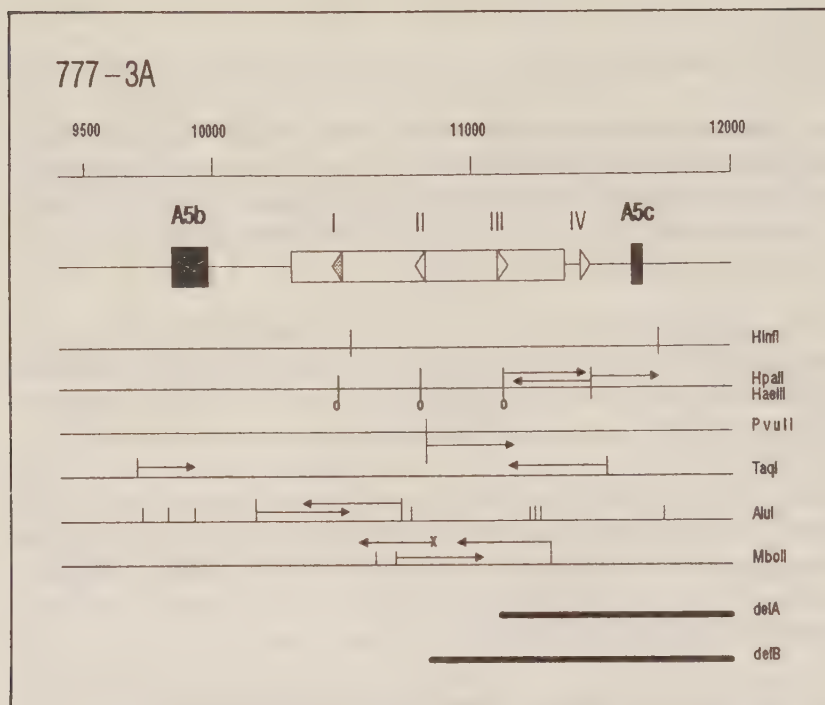


Abb.5 Intron aI5b-Region: Genomstruktur und Sequenzierungsstrategie

Exon A5b und A5c des OXI3-Gens sind durch dunkle, der ORF des Introns aI5b durch den hellen Balken dargestellt. Die Positionsangaben in der oberen Zeile beziehen sich auf die erste Base in Exon A1. Dreiecke geben die GC-Cluster und deren Orientierung wieder (hell dargestellte Cluster sind im Stamm KL14-4A nicht vorhanden). Auf den darunterliegenden Linien sind die Schnittstellen einiger Restriktionsenzyme vermerkt. Mit 'o' markierte Restriktionsschnittstellen fehlen in KL14-4A. Die Pfeile geben Strategie und Umfang der Sequenzanalyse wieder (siehe Text).

3.1.2.1 BESTIMMUNG DER WT-SEQUENZ IM STAMM 777-3A

Die Sequenzierungsstrategie ist aus Abb.5 ersichtlich. Die angegebenen DNA Fragmente wurden durch sekundäre Restriktion der mtDNA gewonnen und wie folgt sequenziert:

HpaII/HpaII- und HpaII/HinfI-Fragmente wurden nach Endmarkierung mit $\alpha^{32}\text{P}$ -dCTP und Strangtrennung entsprechend der chemischen Methode von Maxam & Gilbert (72) sequenziert. Bei den übrigen Fragmenten wurde die M13-Methode (75, 97) angewendet. PvuII/TaqI Fragmente wurden in AccI/SmaI-Schnittstellen von mp8 und mp9 kloniert. AluI/AluI-Fragmente wurden direkt, MboII/MboII-Fragmente erst nach Abbau der 3'überstehenden Enden mit S1-Nuklease und anschließender Auffüllreaktion mit Klenow-Polymerase glatt in die SmaI-Schnittstelle von mp8 kloniert. Vermutlich durch die Inkubation mit S1-Nuklease bedingt, entstanden dabei auch Klone mit verkürzten Inserts, sie wurden zum Teil ebenfalls zur Sequenzierung herangezogen (Fragment 'x' in Abb.5). Die Fragmente wurden von beiden Seiten sequenziert.

Palindrome G+C-reiche Sequenzen (GC-Cluster) führen zu stabilen Rückfaltungen des DNA-Einzelstranges, wodurch die elektrophoretische Mobilität des Fragmentes verändert wird. Die betroffenen Sequenzbereiche werden von herkömmlichen Harnstoffgelen nicht ausreichend denaturiert, wodurch die eindeutige Interpretation des Sequenzgels an den palindromen Bereichen verhindert wird. Daher mußten beheizte Gel-Apparaturen benutzt werden. In einzelnen Fällen wurde es zusätzlich nötig, dGTP durch das Analogon dITP zu ersetzen.

Abb.6 zeigt die Nukleotidsequenz des Stammes 777-3A im Vergleich zu KL14-4A. Wie erwartet entsprechen die drei nur im Stamm 777-3A vorhandenen HpaII/HaeIII-Restriktionsschnittstellen zusätzlichen GC-Clustern. Abgesehen von diesen optionalen Clustern (II-IV) unterscheiden sich die beiden Hefestämme nur durch einzelne Basenaustausche an vier Sequenzpositionen (Pos. 10166; 10811; 10917; 11180), wobei die Proteinsequenz von keiner dieser Abweichungen verändert wird. Die Transition in Pos. 10917 bewirkt in 777-3A die Veränderung einer BglII- bzw. MboI-Schnittstelle in eine HindIV-Schnittstelle. Die aus der Nukleotidsequenz ableitbare Proteinsequenz enthält die für Maturasen typischen Domänen (50). Diese charakteristischen Aminosäurenböcke sind in Abb.6 mit Blk. 1-6 bezeichnet.

Exon A5b 9960 Intron aI5b
k .ATTCCACGATA TTAATTTAATAAGTGTCTGCTTAAAAATTCCTAAATAATATATAA

10007
k TAAATTATAATAATATATAAAAAAAAAATAAAAAAAAAATAAAATAAATTAATATCTTATG

10067
k ATTAATTTTATATAAATAAAAAATTTATTAATATATTATTGGTTATATATATATATATTA

10127 10166
k ATAATAAAAAAAAAATATATATATATATAGCTAACGGGGtttCTCTTATAATTATTATTTA
a ...AGCTAACGGGGGaaCTCTTATAATTATTATTTA
AluI

10187
k TATAATAAATAAGACAATCCCGTGATAACTTTAATATATATATATTATATATTAAAGTAT
a TATAATAAATAAGACAATCCCGTGATAACTTTAATATATATATATTATATATTAAAGTAT

10247
k TGTAGAGACTAAACGTGAATGATTTTAATATTATTTAAATATTAATAATTAAGAGATAGTC
a TGTAGAGACTAAACGTGAATGATTTTAATATTATTTAAATATTAATAATTAAGAGATAGTC
DraI

10307 *** Y Q K⁺ K⁺ N N I I L T T
k CAATCTTATATGTAATATAAGTTAATACCAAAAAAAAAATAATATTATTTTGACTTATT
a CAATCTTATATGTAATATAAGTTAATACCAAAAAAAAAATAATATTATTTTGACTTATT

10367
k T I L I L L I I I L T N N K⁺ V F I E⁻ T L
a ATATATTAATATTATTAATAATAATTTTAACTAATAATAAGTTTTTATAGAACTTTAT
a ATATATTAATATTATTAATAATAATTTTAACTAATAATAAGTTTTTATAGAACTTTAT

10427
k T T T L I F N F Q L I S P F G V P V P G
a ATTATTATTTAATATTTAATTTTCAATTAATATCTCCTTTTGGGGTTCCGGTCCCTGGTC
a ATTATTATTTAATATTTAATTTTCAATTAATATCTCCTTTTGGGGTTCCGGTCCCTGGTC
<===== Cluster I =====

10487
k P A P E⁻ T K⁺ D⁻ I K⁺ N L T E⁻ S I I N N T I
a CGGCCCCCGAAACTAAAGATATTAAGAATTTATATGAATCAATTATAAATAATTATATTA
a CGGCCCCCGAAACTAAAGATATTAAGAATTTATATGAATCAATTATAAATAATTATATTA
===== Hinfi

10547
k N I L N K⁺ T T I N I N K⁺ D⁻ N I N K⁺ L K⁺ F
a ATATTTTAAATAAATATCTTATTAATATTAATAAAGATAATATTAATAAATTTAAATTTT
a ATATTTTAAATAAATATCTTATTAATATTAATAAAGATAATATTAATAAATTTAAATTTT
DraI

10607 Blk.1
k L D⁻ N T T E⁻ E⁻ E⁻ K⁺ G T T (L S G L F E⁻ G D⁻
a TAGATAATTATACTGAAGAAGAAAAAGGTTATTATTATCTGGATTATTGGAAGGAGATG
a TAGATAATTATACTGAAGAAGAAAAAGGTTATTATTATCTGGATTATTGGAAGGAGATG
MboII

10667 Blk.3
 G) N I T T R⁺ C F S I T F (S L E⁻ D⁻ V L) L A
 k GTAATATTTTACTAGATGTTTTTCAATTACTTTTTCTTTAGAAAGATGTTTTATTAGCTA
 a GTAATATTTTACTAGATGTTTTTCAATTACTTTTTCTTTAGAAGATGTTTTATTAGCTA
 MboII AluI

10727 Blk.4
 N T L C T T (F K⁺ I G H⁺ I T A K⁺ T N)
 k ATTATTATGTCTTTATTTTAAAATTGGTCATATTACAGCTAAATATAATT-----
 a ATTATTATGTCTTTATTTTAAAATTGGTCATATTACAGCTAAATATAATTCTCCTCCGG
 DraI AluI N S P P
 <=====

10787 F N K⁺ E⁻ L T A V K⁺ W N I
 k -----tTAATAAAGAATTAACAGCTGTTAAATGAAATATTA
 a CGTCCGCCCCGCGGGCGGGCGGAcTAATAAAGAATTAACAGCTGTTAAATGAAATATTA
 AluI>< del B site
 A S A P R⁺ A G R⁺ T N
 ===== Cluster II ===== PvuII

10847 10860
 I K⁺ K⁺ K⁺ E⁻ Q E⁻ V F I N T I N G K⁺ L L T T
 k TAAAAAAAAGATCAAGAAGTATTTATAAATTATATTAATGGTAAATTATTAACATATA
 a TAAAAAAAAGAGCAAGAAGTATTTATAAATTATATTAATGGTAAATTATTAACATATA

10907 10917 Blk.5
 K⁺ R⁺ T D⁻ Q T F K⁺ T N F N N R⁺ (L N I K⁺ L L
 k AAAGATATGATCAATATTTTAAATATAATTTTAATAATCGTTTAAATATTAAATTATTA
 a AAAGATATGAcCAATATTTTAAATATAATTTTAATAATCGTTTAAATATTAAATTATTA
 (BglII) DraI

10967 Blk.6
 K⁺ P K⁺ E⁻ F D) L T L N P W (L T G F N D⁻ A D⁻
 k AACCTAAAGAATTTGATTACTATTAATCCTTGATTAACAGGTTTTAATGATGCTGATG
 a AACCTAAAGAATTTGATTACTATTAATCCTTGATTAACAGGTTTTAATGATGCTGATG

11027
 G) T F T T G F Q K⁺ H⁺ K⁺ N S Q W L K⁺ F H⁺ L
 k GTTATTTTATCTAGGTTTTCAAAAACATAAAAATAGTCAATGATTAAAAATTCATTAG
 a GTTATTTTATCTAGGTTTTCAAAAACATAAAAATAGTCAATGATTAAAAATTCATTAG

11087 >..... del A site<
 E⁻ L S Q K⁺ D⁻ S T I L
 k AATTATCACAAAAAGATAGTTATATTTTAG-----
 a AATTATCACAAAAAGATAGTTATATTTTAGTCCGCCCCCCCCCGGGGCGGACCCCAA
 L V R⁺ P A P A G R⁺ T P
 ===== Cluster III =====

11147 11180
 D⁻ I I K⁺ K⁺ T F K⁺ T G G I L K⁺ R⁺ D⁻ T K⁺ S
 k -----ATATTATTA AAAAATATTTTAACTTGGTGGTATTTTAAAAAGAGATTATAAATC
 a AGGAGATATTTATTA AAAAATATTTTAACTTGGAGGTATTTTAAAAAGAGATTATAAATC
 K⁺ E⁻ I L L K⁺ N I L N L E⁻ V F ***
 =====> DraI

11207

G A T A T I T K⁺ A Q S S K⁺ A I K⁺ P F I E⁻
 k TGGTGCTACAGCTTATATTTATAAAGCTCAATCATCAAAGCTATAAACCTTTTATTGA
 a TGGTGCTACAGCTTATATTTATAAAGCTCAATCATCAAAGCTATAAACCTTTTATTGA
 AluI

11267

T F N N T Q P L S T R⁺ R⁺ T K⁺ Q T L L L N
 k ATATTTTAATAATTATCAACCATTAAAGTCTTAGAAGATATAACAATATTTATTATTAAA
 a ATATTTTAATAATTATCAACCATTAAAGTCTTAGAAGATATAACAATATTTATTATTAAA
 MboII

11327

I A Y L L K⁺ L N K⁺ L H⁺ I L T N S L L I L
 k TATTGCTTACTTATTAATAATAAATAAATTACATATATTACTTAATTCCTTTATTAATATT
 a TATTGCTTACTTATTAATAATAAATAAATTACATATATTACTTAATTCCTTTATTAATATT

11387

K⁺ N ***
 k AAAGAATTAATATTATTACAAAGTGTTAAAAATATATCTTTAG-----
 a AAAGAATTAATATTATTACAAAGTGTTAAAAATATATCTTTAGTTCCGGGCCCGGTCCAG
 ===== Cluster IV ===

11447

k -----AAATAAAAAATGAATTAATAATAGAGTTAAAAATTATT
 a GGACCGGAACCCGAGAGGAGAAAAATAAATGAATTAATAATAGAGTTAAAAATTATT
 =====>

11507

k ATTAATAAACTTCATTATAACAATATCGAATAATGATAATATTAAAGAGTAAAAATCTTA
 a ATTAATAAACTTCATTATAACAATATCGAATAATGATAATATTAAAGAGTAAAAATCTTA
 TaqI

11567

11603 Exon A5c
 k AAGTGTTAATTAATAATATTCTTTTTTTTTTATGA CTTACTACGTC...
 a AAGTGTTAATTAATAATATTCTTTTTTTTTTATGA CTTACTACGTC...

Abb.6 Intron a15b Region: Nukleotidsequenz

k = KL14-4A (50), a = 777-3A (diese Arbeit). Die Sequenz des nicht-transkribierten Stranges ist gezeigt. Positionsangaben beziehen sich auf das erste Nucleotid von Exon 1. In beiden Stämmen abweichende Nukleotide sind durch Kleinbuchstaben und Positionsangabe gekennzeichnet. Exonanteile sind unterstrichen. GC-Clustersequenzen sind hervorgehoben. Pfeile geben Länge und Orientierung der GC-Cluster I-IV wieder. Die Aminosäuresequenz des ORFs entspricht den Translationsregeln von Bonitz et al. (16) mit Ausnahme von AUA = M (55). Geladene Aminosäuren sind mit ⁺ bzw. ⁻ vermerkt. Der Aminosäurecode steht über der mittleren Base des Triplets. '***' = Stoppcodon. 'Blk.' bezeichnet in Maturasen konservierte Aminosäureblöcke (50). Block 2 fehlt in a15b. '>' '<' begrenzt Klasse A und Klasse B Deletionen.

3.1.2.2 GC-CLUSTER IN aI5b

Das Intron besitzt vier GC-Cluster. Während Cluster I in beiden Stämmen vorkommt, sind die Cluster II-IV auf 777-3A beschränkt. Cluster I-III fallen in den ORF, Cluster IV schließt an diesen an. Die beiden ersten Cluster stören den Leserahmen nicht, Cluster III bewirkt dagegen eine +1 Verschiebung des Leserahmens und damit den Verlust der letzten 68 Aminosäurereste des Proteins. Da jedoch nur Cluster II innerhalb der bei Maturasen konservierten Proteinsequenz liegt (Blk. 1-6), könnte die mögliche Funktion einer aI5b-Maturase erhalten bleiben. Bemerkenswert ist dabei, daß dieses Cluster exakt an einen konservierten Aminosäurenblock (Blk.4) anschließt, welcher sich bei Maturasen durch seine basischen und helixbrechenden Eigenschaften auszeichnet (50). Diese Eigenschaft wird durch das Cluster verstärkt, vier helixbrechenden Aminosäuren (3 x Prolin, 1 x Glycin) werden im Falle der aI5b-Maturase durch das GC-Cluster bereitgestellt.

Während Cluster III-IV in der (5')TAGT-->AAGGAG(3') Richtung (nicht-transkribierter Strang) vorliegen, sind Cluster I-II entgegengesetzt orientiert. Anstelle der optionalen Cluster III-IV findet sich bei KL14-4A (GC⁻-Sequenz) das Trinukleotid TAG. Bei dem optionalen Cluster IV ergibt sich das entsprechende Trinukleotid, wegen der entgegengesetzten Orientierung, aus der Sequenz des komplementären (translatierten) Stranges als TAA. Die abweichende dritte Base des Trinukleotids unterscheidet sich in 777-3A (GC⁺-Sequenz), hier wird ebenfalls ein TAG angetroffen. Diese Beobachtungen führten zu einer Hypothese, wonach das Trinukleotid TAG die bevorzugte Zielsequenz der GC-Cluster-Insertion darstellt (vgl. Targetsite-Hypothese, in Abschn. 3.2.1.2).

3.1.3 OXI3-LEADER

Die OXI3-Leadersequenz lag bisher nicht vollständig vor. Während der 5'Teil des Leaders in der mtDNA des Stammes KL14-4A sequenziert wurde (89), stammen die Sequenzdaten des 3'Teils aus D273-10B (16). Zwischen beiden Sequenzteilen klaffte eine Lücke von 100 bis 150 bp. Da der 5'Deletionsendpunkt der mit-delA-Mutanten diese Sequenzlücke betrifft, wurde der entsprechende Genbereich im wt 777-3A, der Stammform der Deletionsmutanten, sequenziert.

Das diese Sequenz enthaltende EcoRV(2400 bp)-Restriktionsfragment wurde durch Southernblot-Hybridisierung genomischer mtDNA identifiziert und in pUC12 kloniert.

Eine von diesem Fragment angefertigte Restriktionskarte ordnete der Sequenzlücke eine HpaII-Schnittstelle zu. Vor dem OXI3-Promotor unterscheiden sich die Restriktionsdaten von 777-3A durch zwei HpaII-Schnittstellen von KL14-4A. Abb.7 gibt den relevanten Abschnitt des EcoRV(2400 bp)-Fragments wieder. Die Sequenz um die distale (5') HpaII-Stelle wurde in der Zwischenzeit publiziert (34). Sie entspricht einem im Stamm D243-2B-R1 optionalen GC-Cluster. Die Vermutung, daß es sich auch bei der zweiten HpaII-Schnittstelle um ein optionales GC-Cluster handeln könnte, wurde durch die Ausdehnung der Sequenzierungsarbeiten auf diesen Bereich bestätigt.

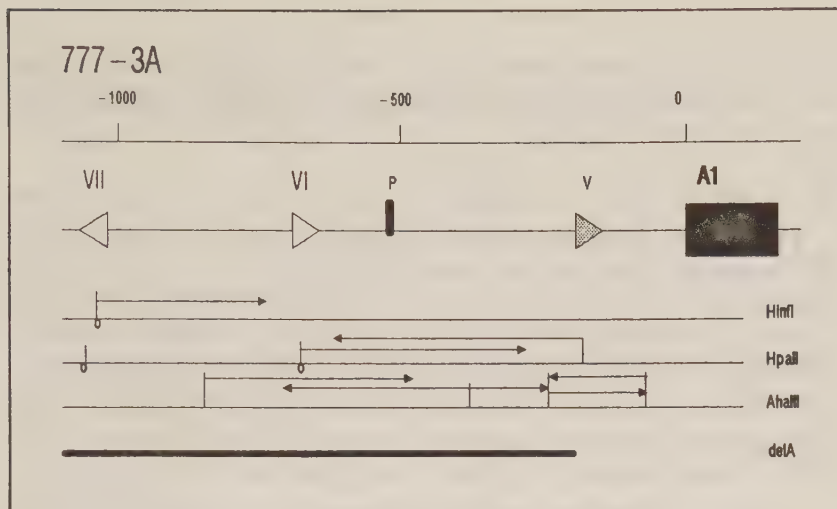


Abb.7 OXI3-Leaderregion: Genomstruktur und Sequenzierungsstrategie

Zur Legende siehe Abb.5. Die Transkriptionsstart Sequenz (Promotor) ist mit 'P' bezeichnet (siehe Text).

3.1.3.1 BESTIMMUNG DER WT-SEQUENZ IM STAMM 777-3A

Die Sequenzierungsstrategie ist aus Abb.7 ersichtlich. Das HinfI-Ende wurde mit $\alpha^{32}\text{P}$ -dATP endmarkiert und nach der chemischen Methode von Maxam & Gilbert sequenziert. Die restlichen Fragmente wurden in M13 subkloniert und entsprechend der Kettenabbruch-Methode (75) sequenziert. Abb.8 zeigt diese Sequenz von 777-3A im Vergleich zu den o.g. Stämmen KL14-4A, D273-10B und D243-2B-R1.

Zwischen D243-2B-3A und 777-3A wurden innerhalb der 140 überlappenden Nukleotide keine Abweichungen festgestellt. KL14-4A und 777-3A unterscheiden sich an zahlreichen Positionen vor dem Nonanukleotid (Pos. 514), welches den vermutlichen Promotor des OXI3-Gens darstellt (89). Diese Sequenzunterschiede beinhalten zwei optionale GC-Cluster, zwei kurze ATATAT-Sequenzen (Pos.-948, Pos.-790), fünf Basenaustausche und 14 Insertionen (Deletionen) ein-

zelner Nukleotide, wobei hierbei 777-3A immer die jeweils längere Form darstellt. Diese Insertionen treten gehäuft in der Umgebung des optionalen Cluster VI auf.

Anschließend an dem wahrscheinlichen Promotor konnten keine Sequenzunterschiede zwischen den beiden Stämmen festgestellt werden. Dies weist darauf hin, daß die Leadersequenz einer stringenteren Kontrolle durch die Evolution unterworfen ist als der davor liegende Spacerbereich, dem wohl nur wenig funktionale Bedeutung zukommt. Innerhalb der 36 überlappenden Nukleotide der beiden Stämme 777-3A und D273-10B konnte jedoch auch innerhalb der Leadersequenz ein Basenaustausch festgestellt werden (Pos.-118). D273-10B dürfte aufgrund seines kurzen Genoms ein geringeres verwandtschaftliches Verhältnis zu den beiden 'langen' Stämmen 777-3A bzw. KL14-4A aufweisen.

3.1.3.2 GC-CLUSTER IM BEREICH DES OXI3-LEADERS

Die in der Abb.8 dargestellten Sequenzen enthalten drei GC-Cluster, wobei die beiden ersten optional, d.h. in KL14-4A nicht vorhanden sind. Die entsprechenden GC-Sequenzen enthalten an Stelle des Clusters wiederum das Trinukleotid TAG bzw. das erweiterte Motiv TAGAG, welches wegen der entgegengesetzten Orientierung des Cluster in seiner komplementären Form (CTCTA) vorliegt. Beide Cluster liegen noch vor der OXI3-Leadersequenz. Das dritte Cluster befindet sich im Leader; wie Restriktionsdaten zeigen, ist die entsprechende HpaII-Schnittstelle auch in den Stämmen D273-10B und KL14-4A vorhanden. Innerhalb dieses Clusters finden die Rekombinationsereignisse statt, welche zu mit-*delA*-Deletionen führen (s.u.).

k ATTAATTATATAAAATATAATA-----CTCTATATT -980
 x ATTAATTATATAAAATATAAATACTCCTTCGGGGTCCCCGCCGGGGCGGGGACTCTATATT
 a ...GACTCTATATT

<===== Cluster VII =====
 HinfI

-946

k AAATATTATTTTTATCAATATTATTATATATATAATAATAATAATAATCAATATT -920
 x AAATATTATTTTTATCAATATTATTAT-----ATAATAATAATAATAATCAATATT
 a AAATATTATTTTTATCAATATTATTAT-----ATAATAATAATAATAATCAATATT

k AATTATTATATATATAAGATTAATATTATTaAATATATTATGAATAATTTAATTAATAA -860
 x AATTATTATATATATAAGATTAATATTATTtAATATATTATGAATAATTTAATTAATAA
 a AATTATTATATATATAAGATTAATATTATTtAATATATTATGAATAATTTAATTAATAA

k ATCTTTAAATATTATCATAAAATATAAAATTAATAaTTCCTTATTATATAAAGAATA -800
 x ATCTTTAAA...
 a ATCTTTAAATATTATCATAAAATATAAAATTAATAaCTCTTATTATATAAAGAATA
 DraI

-790

k ATAATATATA-----ATAATAAGAATGTAAATAATATATATATAATATAATAAAAA -740
 a ATAATATATATATATAATAATAAGAATGTAAATAATATATATATAATATAATAATAAAAA

k ATATATATATATATAAAATATATATATAATATATAG-TAATAATATTTT-ATAT-A -680
 a ATATATATATATATAAAATATATATATAATATATAGATAATAATATTTTTTATATTA

k ATT-A--TTAT-A-TAG-----TA- -620
 a ATTTATTTTATTAATAaTTCGGGGCCCGGCCCGGGAGCCGGAACCCGAAAGGAGTAA
 ===== Cluster VI =====>

k TGA-TA-TATAAA-TCAATATATTAAATAATATATTATATTAGTTCG-TtAgTt-GTA -560
 a TGAATAATATAAAATCAATATATTAAATAATATATTATATTAGTTCGGTcTaaTTGGTA

-514

k TTTTGTAATGAGTAAAA-GTAATATATAATATTAAATAATAAGTATTGATATAAGTAATA -500
 a TTTTGTAATGAGTAAAAAGTAATATATAATATTAAATAATAAGTATTGATATAAGTAATA
 Promotor

k GATATAATAATAATATTATTAATATTTTATATAAAATATTAATAATATAGATTATGAA -440
 a GATATAATAATAATATTATTAATATTTTATATAAAATATTAATAATATAGATTATGAA

k AGAGAGTATTAATATCATTAATATATATATATGTTATATAATTTAAATGATTTTAAATAT -380
 a AGAGAGTATTAATATCATTAATATATATATATGTTATATAAATTTAAATGATTTTAAATAT
 DraI

k ATATATATATATTATATTATAGATTATGATACATTTATATAAAATAATATATATATAAAAA -320
 a ATATATATATATTATATTATAGATTATGATACATTTATATAAAATAATATATATATAAAAA

```

k TTAATTATACTATTACTTTATAATATAATAATATTTATTTATAAAGATATAAAAnAATTG -260
a TTAATTATACTATTACTTTATAATATAATAATATTTATTTATAAAGATATAAAGAATTG

                                >..... del A site .....<
k TTAAAGTTATAACTAAAATATTATATA....Hpa II
a TTAAAGTTATAACTAAAATATTATATAGTCCGGCCCGCCCCCGCGGGCGGACCCCTTT -200
  DraI
                                ===== Cluster V =====

a TTAAGAAGGAGTATTCATTAATAATAAATATTATAAATTCAACTATTGTTATATTTATAA -140
  =====>

                                -118
d Hpa II ....TTATTATCcTTTAAGATATAACAATAATTATTTAAATTAAATTAAAT -80
a ATAGAATAATATATTATTATCaTTTAAGATATAACAATAATTATTTAAA...
                                DraI

d TAAATTTAATTAATTTTTTTTTTTTAAATGAATATAATAATAATAATATTATTAAATTAAT -20

                                Exon A1
d ATATAAAAAAAAAAGTAAAA ATGGTACAAAGATGA...

```

Abb.8 OX13 Leaderregion: Nukleotidsequenz

k = KL14-4A (89), d = D273-10B (16), x = D243-2B-R1 (34), a = 777-3A (diese Arbeit). Positionsangaben beziehen sich auf das erste Nukleotid von Exon 1. In den Stämmen abweichende Nukleotide sind durch Kleinbuchstaben dargestellt. Exonsequenzen sind unterstrichen. GC-Clustersequenzen sind hervorgehoben; Pfeile geben Länge und Orientierung der GC-Cluster V-VII wieder. '>...<' begrenzt Klasse A-Deletionsenden (siehe Text).

3.1.4 DELETIONSKLASSE A

Wie in Abschnitt 3.1 bereits erwähnt, zeichnen sich 22 der 60 OX13-Deletionsmutanten (6% aller mit- des Genoms und damit 0,1% aller Mn^{2+} behandelten Zellen (R. J. Schweyen pers. Mittlg.)) durch die Deletion des gleichen DNA-Bereiches aus. Sie wurden zur Deletionsklasse A zusammengefaßt.

Um die molekulare Struktur des Deletions-Hotspots der Klasse A zu studieren, wurden die Nahtstellen der beiden mit-delA-Mutanten 3001 und MF37 exemplarisch sequenziert. An beiden Mutanten war erwartungsgemäß die gleiche Deletion aufgetreten.

3.1.4.1 SEQUENZANALYSE

Das im wt nicht existierende HinfI-Junktionfragment der beiden mit-delA-Mutanten MF37 und 3001 besitzt eine Länge von ca 1450 bp (vgl. Abb.9).

Das HinfI-Junktionfragment wurde zunächst mit HpaII in vier kleinere Fragmente gespalten. Alle vier Fragmente wurden mit $\alpha^{32}P$ -dCTP endmarkiert und nach Strangtrennung sequenziert (Maxam & Gilbert-Methode). Der Vergleich mit den wt-Sequenzen ergab, daß zwei der Fragmente im Leader, die beiden restlichen in al5b vollständig enthalten waren (vgl. Abb.9). Die Nahtstelle der Deletion wurde offensichtlich durch HpaII gespalten; daher mußte eine andere Sequenzierungsstrategie angewendet werden.

Das HinfI-Junktionfragment wurde daher mit DraI sekundär restringiert, und das Fragmentgemisch (glatt) in die SmaI-Schnittstelle von M13 mp18 kloniert. Klone, welche in der Koloniehybridisierung sowohl mit einem EcoRV(2400 bp)-Fragment (OX13-Leader), als auch mit dem HpaII(300 bp)-Fragment, welches an die Nahtstelle anschließt ('*' in Abb.9) hybridisierten, mußten die Nahtstelle der Deletion enthalten. Sie wurden entsprechend der M13-Methode sequenziert. Der Vergleich mit den ebenfalls sequenzierten wt-Fragmenten (vgl. Abschnitt 3.1.2.1 und 3.1.3.1) erlaubte die Bestimmung der Deletionsstellen.

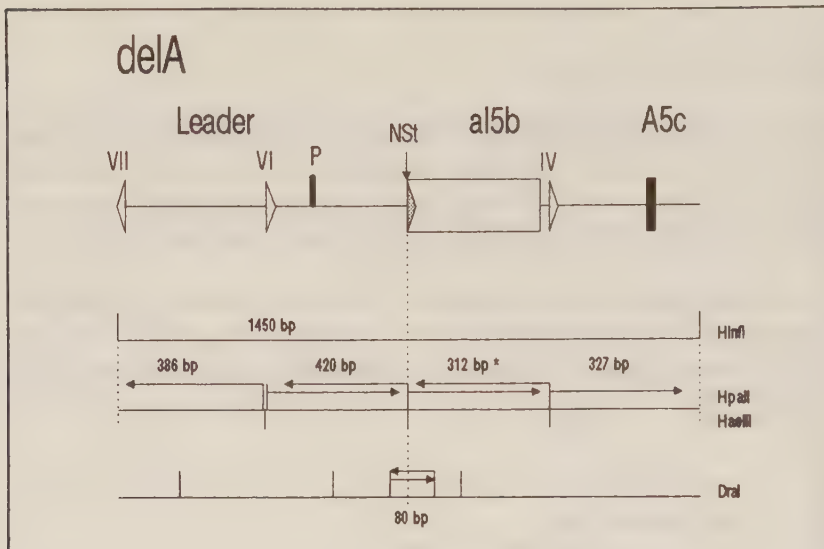


Abb.9 mit-*delA*: Genomstruktur und Sequenzierungsstrategie

Zur Legende vgl. Abb.5. Die Nahtstelle der Deletion ist mit 'NST' gekennzeichnet, sie entstand durch Fusion der Cluster V (vgl. Abb.7) und III (vgl. Abb.5). Das mit '*' markierte Restriktionsfragment wurde auch als Hybridisierungsprobe benutzt (siehe Text).

Abb.10 vergleicht die gemeinsame Sequenz an der Nahtstelle der mit-*delA* 3001 und MF37 mit den entsprechenden wt-Sequenzen. An beiden Seiten der Nahtstelle konnten je etwa 600 bp mit der wt-Sequenz verglichen werden: es traten keine Sequenzunterschiede auf.

Die beiden Deletionsenden werden bei beiden Mutanten durch GC-Cluster gebildet. Beginnend mit den Positionen -232 bzw. +11115 sind die ersten 31 bp der beiden an der Deletion beteiligten Cluster V und III gleich, wobei das in den Deletionsmutanten resultierende Cluster mit Cluster III der rechten Deletionsstelle identisch ist. Daraus ist zu schließen, daß die Deletionsmutanten durch Rekombination innerhalb dieser 31 bp langen, G+C-reichen Homologie entstanden sind.

wt Leader	-232 (Cluster V)	
<u>AAAGTTATAACTAAAATATTATA</u>	<u>TAGTCCGGCCCCGCCCCGCGGGCGGACCCC</u>	<u>tttttaagaaggagtattcattaataat</u>
wt a15b	11115 (Cluster III)	
<u>atcacaaaaagatagttatat</u>	<u>TAGTCCGGCCCCGCCCCGCGGGCGGACCCC</u>	<u>AAAGGAGATATTATTAATAATATTTA</u>
mit ⁻ delA 3001, MF37		
<u>AAAGTTATAACTAAAATATTATA</u>	<u>TAGTCCGGCCCCGCCCCGCGGGCGGACCCC</u>	<u>AAAGGAGATATTATTAATAATATTTA</u>

Abb.10 mit⁻delA-Nahtstelle: Sequenz im Vergleich mit dem wt

Die in den Deletionsmutanten zurückbehaltenen Nukleotide sind in unterstrichen Großbuchstaben, GC-Clustersequenzen fett dargestellt. Der Bereich des Deletionsereignisses ist umrahmt. Die vollständigen Sequenzen wurden in Abb.6 bzw. Abb.8 wiedergegeben.

3.1.4.2 MÖGLICHE DELETIONSMECHANISMEN

Drei unterschiedliche Mechanismen könnten zur mit-delA-Deletion führen:

(i) **Intrachromosomale homologe Rekombination** der beiden Bereiche würde zu den beobachteten Deletionen führen. Da homologe Rekombination vorwiegend als reziprokes Ereignis auftritt, sollten zwei DNA-Ringmoleküle daraus hervorgehen. Während einer der beiden Ringe das Genom der beobachteten mit⁻Mutanten darstellt, entspräche der zweite einem rho-Monomer, dessen Reiteration zu einem rho-Genom führen würde. Tatsächlich konnte gezeigt werden, daß die Exzision von rho-Genomen bevorzugt an GC-Clustern auftritt (4, 127). Nach Segregation der mit/rho-Mischgenome sollten die entsprechenden rho- und mit-Mutanten daher mit vergleichbarer Häufigkeit auftreten. Diese Beobachtung wurde im Falle der OXI3-Mutanten nicht gemacht. Die Diskrepanz könnte erklärt werden durch einen erheblichen Selektionsnachteil des rho-Genoms, welches keinen Startsequenzen der mtDNA Replikation (Origin) besitzen würde. Berichte von rho-Mutanten, die, ohne Originsequenzen zu besitzen, bei Kreuzungsexperimenten dennoch erhebliche Suppressivität bewiesen (33), zeigen, daß die Fitness einzelner Genome auch von anderen Faktoren beeinflusst wird.

(ii) **Slipped Mismatching** während der Replikation der mtDNA würde ebenso zu den beobachteten mit-delA-Mutanten führen. Die Beobachtung, daß der Großteil aller in E.coli auftretenden Deletionen an kurzen direkten Sequenzwiederholungen stattfinden, führte 1982 zu einer Konkretisierung dieses Modells (1). Albertini et al. konnten zeigen, daß die Häufigkeit mit der Deletionen an 4-8 bp langen Sequenzwiederholungen in E.coli auftreten, sowohl der Länge als auch der Entfernung des homologen Bereichs proportional ist, wobei große Entfernungen die Deletionshäufigkeit reduzieren. Das Modell gründet auf einer möglichen Fehlpaarung der während der Replikation aufgeschmolzenen DNA-Einzelstränge an den direkten Homologen, wodurch die dazwischenliegenden DNA-Bereiche ausschleifen würden. Wenn die Replikationsenzyme diesen ausgeschleiften Bereich überspringen können, ist die Deletion fixiert. Alternativ könnte der ausschleifende Einzelstrang vorher durch Reparaturprozesse eliminiert werden. (Ein ähnliches Modell wurde zwei Jahre früher von Efstratiadis et al. entworfen (32)).

Trotz der ausgeprägten Homologie der mit-delA Deletionsendpunkte sollte deren große Entfernung (11347 bp) Fehlpaarungen gemäß dem o.g. Modell erschweren. Zwar können beide Homologe in vivo eng benachbart vorliegen. Problematisch erscheint jedoch die im Modell aufgestellte Forderung, daß beide Homologe Bestandteil des aufgeschmolzenen Bereichs einer einzigen Replikationsgabel sein müssen. Sowohl die vier in al5b lokalisierten GC-Cluster, als auch die eng verwandten Intron al1 und al2 bilden viele zueinander homologe Bereiche in erheblich kürzerem Abstand. Die von daher zu erwartenden häufigen kurzen Deletionen wurden nicht gefunden. Die enorme Häufigkeit der Klasse A-Deletion, welche etwa 6% der mit- des Hefegenoms und 37% der im OXI3 Gen beobachteten Deletionen stellt, läßt m.E. daher auf einen spezifischeren Mechanismus schließen.

(iii) **Transposition** von GC-Clustern hat wahrscheinlich zu deren Ausbreitung im mt Genom geführt. Analogien von GC-Clustern und prokaryontischen IS-Elementen, sowie konservative und replikative Modelle zur Intra- und Interstrang-Transposition werden später diskutiert (vgl. Abschn. 3.2.6 bis 3.2.7). Inwieweit Transpositionsprozesse zu Deletionen führen können, wird dabei erörtert (Abschn. 3.2.7).

3.1.5 DELETIONSKLASSE B

24 der 60 untersuchten mit-Deletionsmutanten gehören der Klasse B an. Mit 40% stellt diese Klasse damit den höchsten Anteil aller OXI3-Deletionsmutanten (vgl. Abschn. 3.1.1). Fünf unabhängig voneinander isolierte mit-delB-Mutanten wurden sequenziert. Bei allen ist erwartungsgemäß die gleiche Deletion aufgetreten.

3.1.5.1 SEQUENZANALYSE

In der Restriktionsanalyse zeichnen sich Klasse B-Mutanten durch ein HpaII/HpaII(3100 bp)- bzw. ein HinfI/HinfI(1350 bp)-Junktionfragment aus. Das leichter isolierbare HpaII-Fragment wurde mit mehreren Enzymen sekundär restringiert. In Abb.11 ist ein HinfI/HpaII(660 bp)-Fragment, welches die Nahtstelle der Deletion enthält, schematisch wiedergegeben.

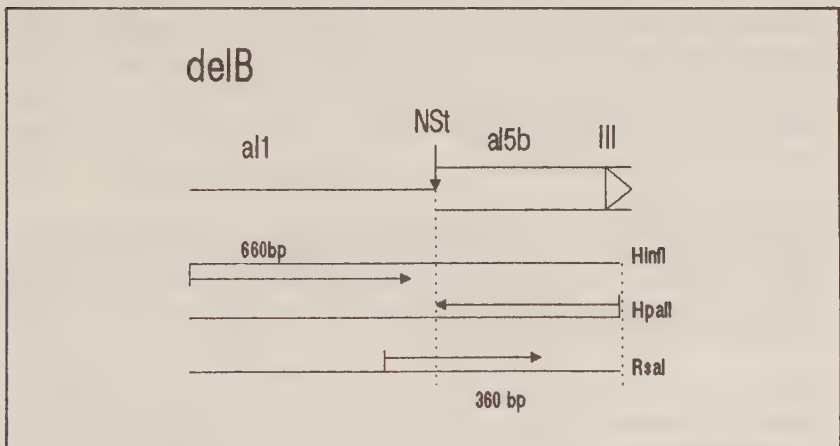


Abb.11 mit-delB: Genomstruktur und Sequenzierungsstrategie

Zur Legende vgl. Abb.5. Die Nahtstelle der Deletion ist mit 'NSt' bezeichnet (siehe Text).

Das *HinfI/HpaII*(660bp)-Fragment der mit-delB 3781 und 7172 wurde mit $\alpha^{32}\text{P}$ -dATP bzw. $\alpha^{32}\text{P}$ -dCTP endmarkiert und nach der chemischen Methode von Maxam & Gilbert von beiden Seiten sequenziert.

Die Sequenzen beider Mutanten waren identisch. Die Nahtstelle der Deletion konnte in der Fragmentmitte lokalisiert werden. Diese Stelle war jedoch zu weit von der Markierungsstelle entfernt, um von den Sequenzgelen eindeutig aufgelöst zu werden.

Die *RsaI/HpaII*(360 bp)-Junktionfragmente der fünf mit-delB-Mutanten 3781, 4021, 7172, 2932, 2613 wurden aus mtDNA präpariert, am 5'Ende mit $\gamma^{32}\text{P}$ -dATP endmarkiert und in ihre Einzelstränge aufgetrennt. Das Laufverhalten der Einzelstrangpaare war bei allen 5 Mutanten identisch (Daten nicht gezeigt).

Nach Sequenzierung der Einzelstränge (Maxam & Gilbert) konnte die Nukleotidsequenz dreier Mutanten eindeutig erhalten werden. Das Bandenmuster in der Sequenzleiter dieser Mutanten (2613, 2932, 3781) entsprach zu beiden Seiten der Deletionsnahtstelle jenem der schlechter interpretierbaren Sequenzleitern der Mutanten 4021 und 7172. Obwohl das Sequenzgel an der Nahtstelle der Deletionsmutanten 4021 und 7172 nicht eindeutig interpretiert werden konnte, muß gefolgert werden, daß bei den beiden Mutanten ebenso viele Nukleotide deletiert sind, wie bei den mit-del 2613, 2932 und 3781. Die Identität aller fünf Mutanten kann daher angenommen werden.

An der rechten Deletionsstelle konnte die delB-Sequenz mit der des ebenfalls sequenzierten wt 777-3A (vgl. Abschn. 3.1.5.1) verglichen werden. An der linken Stelle fehlte die 777-3A-Sequenz; deshalb wurde der Vergleich delB(777-3A)/wt(D273-10B) herangezogen.

Insgesamt wurden zu beiden Seiten der Nahtstelle je ca. 300 bp mit den entsprechenden wt-Sequenzen verglichen: es traten keine Sequenzunterschiede auf. Aus dem Vergleich der Sequenz an der Nahtstelle der Deletion mit den entsprechenden wt-Sequenzen konnten die Deletionsorte eindeutig festgelegt werden (vgl. Abb.12).

		2619		
wt D273-10B		Intron aI1 Exon A2		
<u>CGTGGCAACATAGGTTAATTGCTATTTCAT</u>		ttttagtagttgggtcatgctgtattaatag		
----->		-->		
wt 777-3A	Intron aI5b	10840	Intron aI5b	
ggactaataaagaattaacagctgttaaagt		<u>AAATATTATAAAAAAAAAAGAACAAGAAGT</u>		
<-----<				
mit ⁻ delB 2613,2932,3781				
<u>CGTGGCAACATAGGTTAATTGCTATTTCAT</u>		<u>AAATATTATAAAAAAAAAAGAACAAGAAGT</u>		

Abb.12 mit⁻delB-Nahstelle: Sequenz im Vergleich mit dem wt.

Die in den Deletionsmutanten zurückbleibenden Nukleotide sind in Großbuchstaben dargestellt und unterstrichen. Die vollständigen Sequenzen sind in Abb.6 und Abb.8 wiedergegeben. Die Positionen der Deletionsendpunkte sind angegeben. Palindrome Sequenzen sind mit Pfeilen unterlegt; sie ermöglichen die in Abb.13 dargestellte Sekundärstruktur (siehe Text).

Die Deletion beginnt am Übergang von Intron aI1 zu Exon A2 mit der ersten Base des Exons (Pos. 2619) und endet 8221 bp weiter (Pos. 10840) im ORF des Introns aI5b.

3.1.5.2 MÖGLICHE DELETIONSMECHANISMEN

Im Gegensatz zu den Klasse A-Mutanten sind linke und rechte Deletionsstelle der Klasse B-Mutanten zueinander nicht homolog. Die auf solchen Homologien basierenden, für Klasse A-Mutanten diskutierten Rekombinationsmechanismen können daher nicht an der etwa gleich häufigen Entstehung der Klasse B-Mutation beteiligt sein.

3.1.5.2.1 PAARUNGSMÖGLICHKEIT DER DELETIONSENDEN

Die beiden Klasse B-Deletionsenden werden von zueinander invertierten Nukleotidsequenzen gebildet (vgl. Abb.12). Intrastrangpaarung könnte daher die Annäherung der beiden Rekombinationsstellen bewirken. In Abb.13 ist eine mögliche Form der Basenpaarung nach der Rückfaltung eines Stranges wiedergeben.

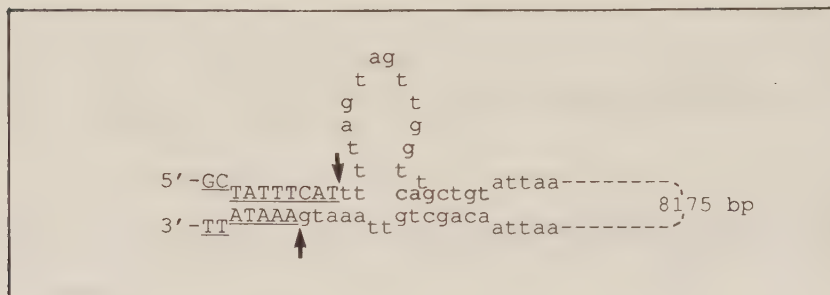


Abb.13 Mögliche Intrastrangpaarung der Klasse B-Deletionsorte

Die in Klasse B-Mutanten deletierten DNA-Bereiche sind in Kleinbuchstaben dargestellt. Pfeile bezeichnen die Deletionsendpunkte.

Die sequenz-spezifische Exzision müßte an den mit Pfeilen gekennzeichneten Stellen stattfinden. Enzyme, welche in kruzförmigen Strukturen superhelikaler DNA erkennen und schneiden, wurden in *Sacch.cerevisiae* beschrieben (114). Da beide Schnittpunkte Exon-Intron-Übergängen entsprechen (s.u), könnten auch Faktoren des Spleißapparates an der Exzision beteiligt sein. Nach der Ligation wäre das Deletionsereignis abgeschlossen.

3.1.5.2.2 KLASSE B-DELETIONSENDEN ENTSPRECHEN SPLEIßPUNKTEN

Auffallend ist der Befund, daß der linke Deletionsort mit dem 3'Spleißpunkt von Intron a11 zusammenfällt. Dies könnte auf einen Zusammenhang der Deletion mit dem Spleißprozeß oder typischen Sequenzen an den Exon-Intron-Grenzen hindeuten. Daher wurde gefragt, ob auch der rechte Deletionsort Ähnlichkeit mit der Spleißsequenz des Intron a11 hat. Abb.14 zeigt die weitgehende Übereinstimmung von rechtem Deletionsort und dem 3'Ende des Exon A1.

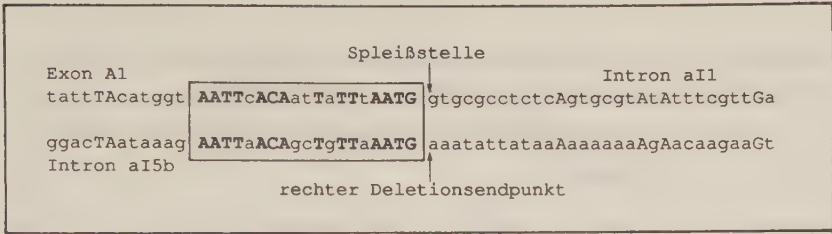


Abb.14 Sequenzhomologie der rechten Deletionsstelle in a15b mit der 5'Spleißsequenz des Introns a11

Der homologe Bereich ist umrahmt. Spleißstelle und Deletionsendpunkt sind mit Pfeilen markiert (siehe Text).

An Spleiß- bzw. Deletionsstelle sind 14 von 19 bp identisch, was einer Homologie von 74% entspricht. Beide Sequenzen interagieren mit dem Übergang von Intron a11 zu Exon A2, was zum Spleißen des Intron a11 bzw. zu den Klasse B-Deletionen führt (vgl. Abb. 15).

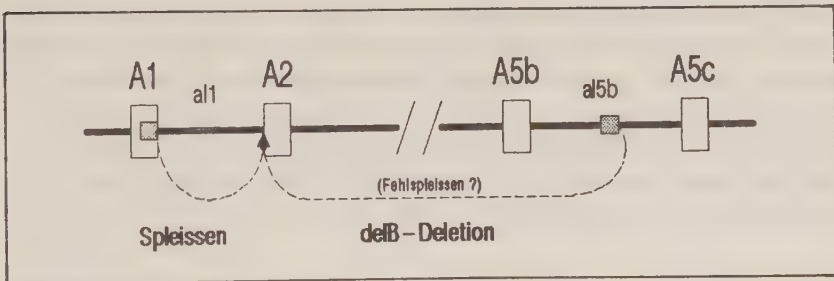


Abb.15 Klasse B-Deletion durch Fehlspleißen

Exons sind durch helle Balken dargestellt. Die homologen Bereiche am Ende von Exon1 bzw. am Ende des deletierten Bereiches sind durch schattierte Blöcke wiedergegeben, ihre Sequenzen wurden in Abb.14 verglichen (siehe Text).

Die Deletionsenden der Klasse B-Mutanten suggerieren damit einen dem Intron Spleißen entsprechenden Deletionsmechanismus, bei dem das Exon A2, statt an das Exons A1 (spleißen) an die dazu homologe Sequenz im Intron a15b

(fehlspleißen) ligiert wird. Auf welche Weise Fehlspleißen zu den mit-delB-Deletionen führen könnte, wird in Abschn. 3.1.5.2.5 näher erläutert.

3.1.5.2.2.1 Trans-Spleißen

Der Spleißmechanismus wird im Rahmen dieser Arbeit nicht ausführlich behandelt (als rezente Übersichtsartikel siehe (44, 90)). Während zunächst die einzelnen Schritte des tRNA-Spleißens verstanden wurden (45), konnten in den beiden letzten Jahren wesentliche Fortschritte im Verständnis der übrigen Hauptklassen des RNA-Spleißens erzielt werden.

Spleißen von Kerngen-Transkripten (hnRNA) ist nur von verhältnismäßig wenigen konservierten Primär- und Sekundärstrukturen innerhalb des Introns abhängig. Stattdessen wird eine Vielzahl von in trans wirkenden Elementen für effizientes, basengenaues Spleißen benötigt. Die wichtigsten Sequenzsignale innerhalb des Introns (5'Ende, Verzweigungspunkt, 3'Ende) dienen zumindest zur korrekten Assemblierung des Spleißapparates (Spleißosom) (90, 44).

Wegen der bemerkenswerten Fähigkeit zumindest einiger Gruppe I- und Gruppe II-Introns, in vitro autokatalytisch zu spleißen (63, 92, 120, 132), ist erkennbar, daß diese Moleküle weniger von aufwendigen Spleißapparaten abhängig sind, als vielmehr von ihrer primären und sekundären Struktur. Tatsächlich wurden Gruppe I- und Gruppe II-Introns zunächst auf Grund ihrer Primär- und Sekundärstrukturen erkannt und unterteilt (78, 79). Neuere Erkenntnisse über ihre Spleißmechanismen bestätigen diese Klassifikation (20).

Es scheint heute, daß sowohl nukleäres hnRNA-Spleißen, als auch das Spleißen von Gruppe I- und Gruppe II-Introns, im wesentlichen aus zwei unabhängigen Transesterifikations-Reaktionen besteht:

(i) Lösen der 5'Exon-Intron-Bindung. Dies geschieht durch einen nukleophilen Angriff am 5'Exon-Intron-Übergang durch eine freie OH-Gruppe. Bei Gruppe I ist die angreifende Gruppe das 3'OH eines freien Guanotin-Moleküls (132, 21), bei Gruppe II und bei Kerngenen ist es das 2'OH eines Nukleotidrestes nahe dem 3'Ende des Introns (Verzweigungspunkt) (120, 92). Letzteres führt zur einer kovalenten Bindung von Verzweigungspunkt und 5'Ende des Introns unter Ausbildung einer lassoförmigen Struktur (Lariat).

(ii) Lösen der Intron(3')-Exon Bindung und Ligation der Exons, durch einen nukleophilen Angriff der OH-Gruppe am 3'Ende des (in der ersten Reaktion freigewordenen) 5'Exons auf den Intron(3')-Exon Übergang. Das Intron wird dabei freigesetzt.

Allen drei Spleißmodi ist gemein, daß das 5'Exon nach dem ersten Reaktionsschritt nicht mehr kovalent gebunden ist. Die Diffusion des 5'Exons muß daher durch andere Faktoren verhindert werden. Im Falle nukleären Spleißens wird das 5'Exon vermutlich durch einen RNA-Protein-Komplex (snRNP, Spleißosom) gebunden, wobei Basenpaarung von Exonende und U1 RNA eine wesentliche Rolle spielen könnte (65). Bei Gruppe I-Introns wird die Fixierung des 5'Exons wahrscheinlich durch eine komplementäre Sequenz (Guidesequenz) innerhalb des Introns bewirkt (27, 76, 122). Die Paarung des 5'Exons mit dieser spezialisierten Intronsequenz wird durch neuere genetische Arbeiten gestützt (20, 123).

Ein Fehlspleißen, bei dem an Stelle des 5'Exons eine dazu homologe Sequenz an das 3'Exon ligiert wird, sollte daher prinzipiell möglich sein. Tatsächlich konnte bei Gruppe I-Introns gezeigt werden, daß Oligonukleotide, welche dem 3'Ende des 5'Exons entsprechen, das 5'Exon in trans ersetzen können. Die zweite Transferifikation resultiert dabei in der Spaltung der Intron(3')-Exon-Bindung und in der Ligation des Oligonukleotids an das 3'Exon (41, 57).

Die Ähnlichkeiten des Spleißprozesses lassen vermuten, daß auch beim Spleißen von Gruppe II ähnliche Paarungen existieren. Bisher konnten jedoch keine Faktoren identifiziert werden, welche das 5'Exon beim Spleißen von Gruppe II Introns festhalten können.

Die Bedeutung der Sequenz am 3'Ende des 5'Exons beim Spleißen von Gruppe II konnte indes von Jacquier und Rosbash bestätigt werden (59). Die Autoren klonierten einen Teil des mitochondrialen OXI3-Gens von *Sacch.cerevisiae* (Stamm KL14-4A) in den sp6 Transkriptionsvektor. Das Plasmid enthielt 52 bp des Exons A5, Intron aI5 und 183 Nukleotide des Exons A6. Eine TaqI Restriktionsschnittstelle am Ende des Exons A5 ermöglichte die vorzeitige Terminierung ('run-off'-Transkription) der in vitro-Transkription, exakt eine Base (Guanin) nach dem Exonende (im Folgenden als '5'Exon-G Transkript' bezeichnet). Durch eine Reihe von in vitro-Spleißversuchen konnten die Autoren zeigen, daß das 5'Exon-G Transkript in trans an das 3'Exon des nicht vorzeitig terminierten (Exon-Intron-

Exon-) Transkripts gespleißt wird. Die Sequenzierung des Spleißproduktes zeigte, daß das terminale Guanosin-Nukleotid des Exon-G-Transkripts, welches der ersten Base des Introns entspricht, vor der Ligation entfernt wurde: die Exons wurden also basengenau ligiert.

Transkripte, welche nur 2 bp des Exons A5 + Intron a5 + 183 bp des Exons A6 enthielten (Intron(3')-Exon), waren nicht in der Lage, in vitro autokatalytisch zu spleißen. Damit ist zu vermuten, daß mehr als 2 bp der 5'Exonsequenz zum Spleißen von Gruppe II-Introns nötig sind. Die gemeinsame Inkubation der Exon-G- und Intron(3')-Exon-Transkripte resultierte wieder in der exakten Ligation der Exons. Die Reaktionsgeschwindigkeit des Trans-Spleißens erreichte diejenige des Cis-Spleißens von langen (Exon-Intron-Exon-) Transkripten.

Damit wird klar, daß Gruppe II-Introns zumindest in vitro zum effektiven Trans-Spleißen befähigt sind. Die einzige Anforderung an die kryptische Spleißstelle besteht in deren Homologie mit dem Ende des originalen 5'Exons. Das in Abb.16 skizzierte Fehlspleißen könnte damit ein häufiges Ereignis bei der Prozessierung des OXI3-Transkripts beschreiben, da Intron a11 ebenfalls der Gruppe II angehört. In Abb.13 wurde gezeigt, daß die Annäherung dieser vermutlich kryptischen Spleißstelle an das Exon A2 durch weitere Paarungsmöglichkeiten erleichtert werden kann.

3.1.5.2.2.2 Spezifität der Deletionsstelle

Die als bedeutungsvoll vermutete Homologie zwischen dem 3'Ende von Exon A1 und dem 3'Deletionsende (Abb.14) warf die Frage auf, ob noch weitere Sequenzen innerhalb des OXI3-Gens Ähnlichkeiten mit dem Ende des Exon 1 aufweisen (Sequenz AATTCACAATTATTTAATG).

Eine Computeranalyse erbrachte keine größere Homologie vergleichbarer Länge innerhalb des OXI3-Bereichs. Es konnten jedoch drei Sequenzpositionen gefunden werden, welche gleichstarke (14 von 19 bp) Homologie zum A1-Ende hatten. Die beiden ersten liegen innerhalb von Intron a12 (Pos.5263, 5863), die dritte betrifft ebenfalls ein Exonende (A6, Pos.12793). Die Existenz des Introns a16 wurde aus dem Vergleich der aus der Nukleotidsequenz abgeleiteten Aminosäuresequenz mit dem entsprechenden Protein des Menschen geschlossen (50). Das nur 21 bp lange Intron entspricht nicht den in Mitochondrien üblichen Introns, welche

auf Grund ihrer Sekundärstrukturen in die beiden Gruppen I und II eingeteilt werden (vgl. Abschn. 3.1.5.2.2.1). Da der Leserahmen durch das Intron nicht unterbrochen wird, kann die Existenz des Introns angezweifelt werden.

Deletionen zwischen der Exon-Intron-Grenze aI1/A2 und den o.g. mit dem Exon 1-Ende homologen Genloci wurden bisher nicht beobachtet. Die in Abb.14 gezeigte Homologie zwischen der rechten Klasse B-Deletionsstelle und dem Ende von Exon A1 scheint damit nicht alleine auszureichen, um die extrem häufige mit-delB-Deletion zu erklären.

Bei einer vergleichbaren Computeranalyse, bei der nach Homologie mit dem 3'Ende des deletierten Bereiches (Sequenz AATTAACAGCTGTAAATG) gesucht wurde, konnte innerhalb des OXI3-Bereiches außer dem Ende von Exon A1 keine weitere Sequenz gefunden werden, in welcher wenigstens 14 der 19 Nukleotide übereinstimmen. Spezifische Positionen innerhalb dieser Sequenz könnten damit von Bedeutung sein.

3.1.5.2.3 HÄUFIGE DELETIONEN AN EXON-INTRON-GRENZEN IN UNTERSCHIEDLICHEN ORGANISMEN

Deletionen an Exon-Intron-Übergängen konnten mittlerweile in unterschiedlichen Organismen als häufiges Ereignis beobachtet werden.

Untersuchungen von spleißdefekten Intronmutanten des COB-Gens von *Sacch.cerevisiae* brachten überraschende Ergebnisse. Die Mutanten revertierten spontan mit einer Häufigkeit von 10^{-7} bis 10^{-9} durch die exakte Deletion des Introns. In vielen Fällen deletierten die Intron bI1, bI2 u. bI3 gleichzeitig, wofür kein offensichtlicher Selektionsdruck vorhanden war (40). Insbesondere die gleichzeitige Deletion der Introns läßt die Beteiligung von maturer mRNA wahrscheinlich erscheinen. Gargouri et. al. diskutierten daher folgende zwei Hypothesen:

(i) **Heteroduplexstrukturen** aus DNA und maturer oder zumindest partiell gespleißter mRNA. Die aus der Heteroduplex ausschleifende intronische DNA könnte dabei von Reparaturprozessen eliminiert werden.

(ii) **Homologe Rekombination** zwischen genomischer DNA und der DNA revers transkribierter mRNA.

Beide Hypothesen implizieren das Vorhandensein von gespleißter RNA in den Mitochondrien der (weitgehend) spleißdefekten Mutanten. Die zweite Hypothese fordert darüber hinaus reverse Transkriptase-Aktivität innerhalb der Mitochondrien von *Sacch.cerevisiae*.

Die Seneszenz einiger filamentöser Pilze steht unter cytoplasmatischer Kontrolle. In seneszenten Mycelien von *Podospora anserina* konnten mt Plasmide gefunden werden, die in jungen Zellen nicht vorhanden sind (112). Diese Plasmide enthalten Tandem-Wiederholungen mitochondrialer DNA. Wie Osiewacz und Esser zeigen konnten, entspricht das dabei am häufigsten auftretende α -SEN-Plasmid (pl-DNA) exakt dem Intron 1 des Cytochrom-Oxidase-UE 1-Gens (88). Intron 1 fehlt dann in den mt Genomen seneszenten *P.anserina* Mycelien (5). Sowohl die o.g. Hypothesen als auch die Möglichkeit einer direkten Intronexzision aus der mtDNA wurden dazu diskutiert (5).

Die o.g. Beispiele zeigen, daß die exakten Exon-Intron-Grenzen bevorzugte Stellen der DNA Deletion darstellen können. An *Cauliflower Mosaic Virus* (CaMV) konnte gezeigt werden, daß solche Deletionen auch an Sequenzen auftreten, welche den Spleißstellen nur ähnlich sind (51). In einem Stamm des Pflanzenvirus, in dessen Lebenszyklus reverse Transkription eine wesentliche Rolle spielt, tauchen sequenz-spezifische Deletionen derart häufig auf, daß die resultierenden Junktionfragmente in den Restriktionsgelen der wt-DNA als minore Banden hervortreten. Die Sequenzanalyse konnte zeigen, daß die beiden Deletionsenden den für Eukaryonten typischen Donor- und Akzeptor-Konsenssequenzen des RNA-Spleißens ähneln (51).

Durch Einführung von gezielten Basenaustauschen in die Donor-Deletionssequenz konnten die Autoren das Auftreten dieser Deletionen unterbinden. Stattdessen traten andere sequenz-spezifische Deletionen auf. Sequenzierung dieser Deletionsnahtstellen zeigte, daß die jeweils nicht mutagenisierte Akzeptor-Deletionssequenz weiterhin an den Deletionseignissen beteiligt blieb, während die neu aktivierten kryptischen Deletionsequenzen wiederum mit den Donor-Konsenssequenzen des RNA-Spleißens homolog war, wenn auch in etwas gerin-

gerem Maße. In anderen CaMV Stämmen wurden die entsprechenden Deletionen nicht oder seltener gefunden.

Zumindest für Retroviren und andere Genome, deren Replikation reverse Transkription beinhaltet, muß daher ein Mechanismus angenommen werden, welcher wahllose reverse Transkription gespleißter RNA verhindert, um die Funktionsfähigkeit des Genoms zu erhalten. Einem solchen Mechanismus könnten fehlgespleißte RNA-Moleküle möglicherweise entgehen.

3.1.5.2.4 REVERSE TRANSKRIPTASE-AKTIVITÄT IN MITOCHONDRIEN

Rekombination von mtDNA mit cDNA prozessierter RNA setzt die Anwesenheit reverser Transkriptase in Mitochondrien voraus. Assoziiert mit Ty-Elementen, konnte diese Aktivität in *Sacch.cerevisiae* nachgewiesen werden (15). Vergleichende Computeranalysen lassen auch in den Mitochondrien die Codierung reverser Transkriptase wahrscheinlich erscheinen (77).

Sowohl die mt ORFs in Intron 1 des Cytochrom-Oxidase-UE 1-Gens in *Sacch.cerevisiae* (aI1), *P.anserina* (α -SEN-Intron) und *Schizosacch.pombe* (cob-Intron 1), als auch das mt Intron aI2 von *Sacch.cerevisiae* und das mt Mauriceville-Plasmid von *N.crassa* zeigen Homologie auf Proteinebene mit reversen Transkriptasen verschiedener transponierbarer Elemente und Viren, wie z.B dem *Cauliflower Mosaic Virus* (CaMV) (77). Durch in vitro-Versuche konnte die Aktivität der reversen Transkriptase während der Seneszenz von *P.anserina* zwischenzeitlich bestätigt werden (113).

Sollte reverse Transkription in Mitochondrien eine wesentliche Rolle spielen, so müßte auch hier ein Mechanismus vorhanden sein, welcher wahllose DNA-Synthese aus RNA verhindert.

Wie oben erwähnt, treten Deletionen des COB-Introns bI1 in den Mitochondrien von *Sacch.cerevisiae* mit großer Häufigkeit (10^{-7} bis 10^{-9}) auf. Durch Einbringen der Mitochondrien des Stammes 777-3A in vier verschiedene Hefestämme (Cytoduktion) konnte gezeigt werden, daß die Exzisionsrate von bI1 vom Kerngenom abhängt (22). So deletierte dieses Intron in den Stämmen AH215 und AH216 etwa 100 mal häufiger als in den Stämmen IC8 und IC11. Die offensichtlich beteiligten Kerngene konnten bisher nicht identifiziert werden. Da reverse Transkription

innerhalb der Mitochondrien aus o.g. Gründen vermutet werden kann, ist m.E. auch in *Sacch.cerevisiae* ein Mechanismus zur Verhinderung der reversen Transkription prozessierter RNA denkbar. Fehler innerhalb eines solchen Mechanismus könnte die unterschiedliche Deletionsrate mitochondrialer Introns in verschiedenen Hefestämmen erklären. Da die möglichen Mechanismen vermutlich zwischen verschiedenen RNA-Molekülen unterscheiden könnte, ließen sich damit auch unterschiedlich häufige Deletionen verschiedener Introns innerhalb eines Hefestamms erklären.

(Anmerkung: Ein bisher wenig beachtetes Phänomen besteht in dem spezifischen Verlust von Emunctator-Mucin der monorrhinen Rhinogradentia Gattung *Columnifax*. Die zur selben Familie (*Rhinocolumnidae*) gehörende Gattung *Dulcicauda* benutzt Emunctator-Mucin, welches eine pentosehaltige Mucotinschwefelsäure enthält, zum Aufbau ihrer Sella (115). Haben sequenz-spezifische Deletion (z.B. von Introns) dazu geführt, daß *Columnifax* Arten heute nicht mehr in der Lage sind, Emunctator-Mucin zu synthetisieren? Die Antwort auf diese schwierige Frage wird uns auch in Zukunft versagt bleiben.)

3.1.5.2.5 MÖGLICHE ROLLE DER RNA BEI MITOCHONDRIALEN DELETIONEN

Die Deletionsenden der in dieser Arbeit untersuchten Klasse B-Mutanten legen die Mithilfe von Spleißmechanismen an der Deletionsausbildung nahe. Die Beteiligung (fehl-) gespleißter (revers transkribierter) RNA erscheint hier jedoch problematisch. Während bei den Deletionen der COB Intron bI1, bI2, bI3, bzw. des α -SEN-Introns und den Deletionen in CaMV jeweils das Intron bzw. der einem Intron entsprechende DNA Bereich deletierte, entspricht der deletierte DNA-Abschnitt der Klasse B den Exonbereichen (vgl. Abb.16.).

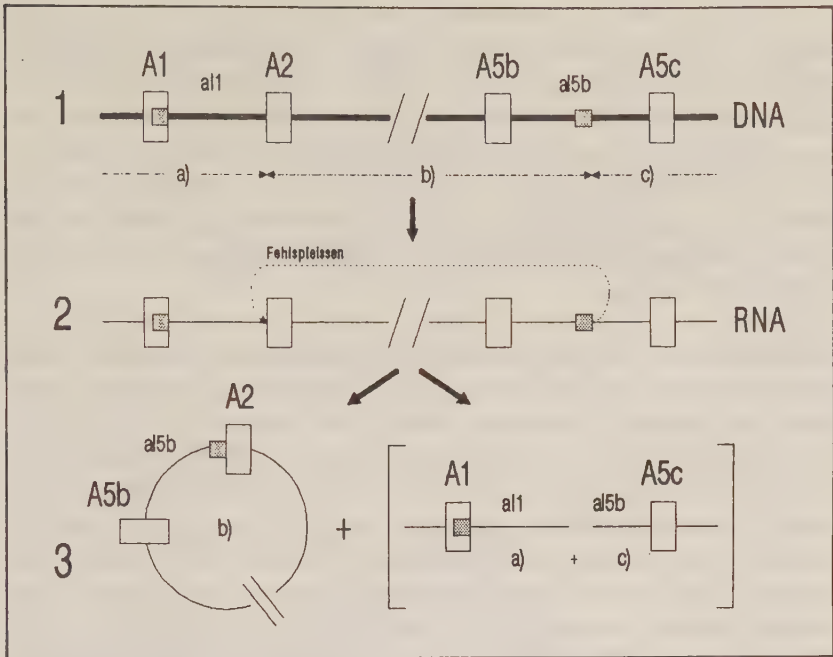


Abb.16 Struktur der mit delB-RNA nach einem hypothetischen Fehlspleißen

Zur Legende vgl. auch Abb15. In Teil 1 der Abb. ist der betroffene Abschnitt des OXI3-Gens wiedergegeben. Unterbrochene Pfeile bezeichnen DNA-Bereich a), b) und c). Deletion des mit 'b)' markierten Bereiches führt zu Klasse B-Mutanten (siehe Text).

Das Exon A2 wird, an Stelle an das Exon A1, an die dazu homologe Sequenz in a5b gespleißt (Teil 2), wodurch es zu einem Ringschluß des Spleißproduktes kommt (Teil 3 b)). Die Ligation des in den Klasse B-Mutanten verbundenen 3'Ende von a11 (a) mit der rechten Hälfte von a15b (b)) ist damit nicht erreicht: sie entspräche einer Ligation von 5'Ende mit dem 3'Ende eines Introns während des RNA-Spleißens.

Nach dem derzeitigen Verständnis des Spleißvorganges werden die beiden Enden des gespleißten Introns nicht ligiert. Stattdessen verbindet sich das 5'Ende des Introns mit einem Nukleotid nahe dem 3'Ende zur bekannten Lariatstruktur (vgl.

Abschn. 3.1.5.2.2.1). Sollen die von Gagouri et al. diskutierten Hypothesen (DNA/RNA Heteroduplex bzw. Rekombination mit revers transkribierter RNA (s.o.)) auf die mit-delB-Deletionen angewendet werden, so muß anstelle der Lariatbildung die Ligation der Intronenden angenommen werden. Dies gilt auch für das α -SEN-Plasmid von *P.anserina*, da das Plasmid aus exakt in tandem reitierter Intron 1-DNA besteht und somit nicht durch reverse Transkription des Lariats entstanden sein kann. In beiden Fällen ist das Intron 1 der Cytochrom-Oxidase-UE 1 beteiligt.

Das Phänomen ist indes nicht auf dieses Intron beschränkt. Im Anschluß an diese Ergebnisse wurden auch in *Schizosacch.pombe* vergleichbare Deletionen festgestellt. Drei unabhängig voneinander isolierte COB Mutanten (EB4, EB7, EB11) weisen identische Deletionen auf. Auch hier beginnt die Deletion mit der ersten Base des Exon 2, sie endet im trailer. Wie bei den mit-delB-Mutanten ist der rechte Deletionsendpunkt mit dem 5'Ende des betroffenen Intron 1 homolog (F.Ahne, in Vorbereitung). Das Ausmaß dieser Homologie ist in beiden Fällen ähnlich (12 von 15 bp in *Schizosacch.pombe*, 14 von 19 bp in *Sacch.cerevisiae*). Beiden Deletionen scheint ein vergleichbarer Mechanismus zugrunde zu liegen. Rekombination mit fehlgespleißter, revers transkribierter RNA scheint in beiden Fällen nicht möglich, da die RNA Enden, welche den in den Deletionsmutanten ligierten DNA Abschnitten entsprechen, beim Spleißprozess nicht verbunden werden (s.o.). Vielmehr würde das 5'Ende an einer 5'-2' Lariatbindung beteiligt sein.

Abb.17 zeigt auf welche Weise eine Heteroduplexstruktur aus genomischer mtDNA und fehlgespleißter RNA die Klasse B-Deletionsenden zusammenführen könnte. Anders als bei den Deletionen des COB Introns von *Sacch.cerevisiae* diskutiert, würde diese RNA nicht mit dem zurückbleibenden, sondern mit dem deletierenden DNA-Bereich paaren. Abb.17 zeigt, daß auch dabei die basengenaue Positionierung der beiden Deletionsenden erreicht würde.

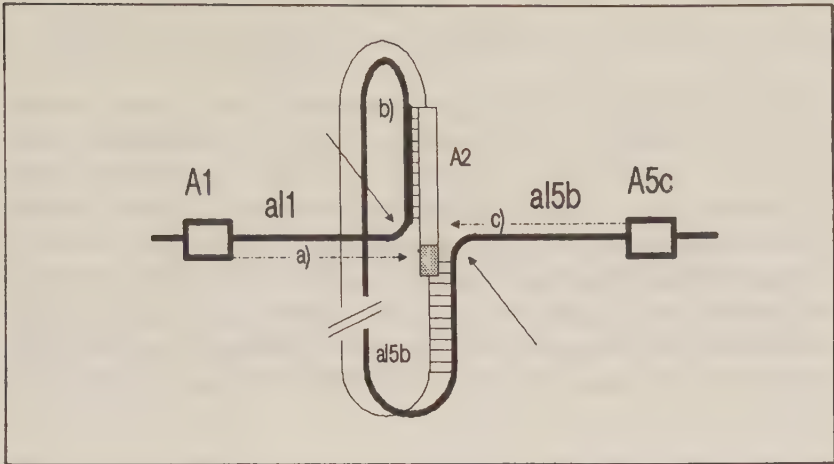


Abb.17 Heteroduplex zwischen mtDNA und fehlgespleißter RNA

Die RNA ist durch dünne Linien dargestellt, sie entspricht der im Teil 3 der Abb.16 mit 'b)' bezeichneten RNA. Die DNA (dicke Linie) kann im deletierenden Bereich b) mit der RNA paaren. Die Deletionsendpunkte sind durchgehende Pfeile gekennzeichnet (vgl. auch Abb.16) (siehe Text).

Schnitte an den Deletionsenden und nachfolgende Ligation von Intron 1 mit dem rechten Teil des Intron a15b sollten dabei von den gleichen Faktoren katalysiert werden, da bei einer Trennung der beiden Prozesse die Deletionsenden nach dem Schnitt möglicherweise diffundieren könnten.

Ein entsprechender Mechanismus wäre auch in der Lage, die exakte Fusion der Intronenden im α -SEN-Plasmid von *P.anserina* zu erklären. Die in der Abb.17 zirkulär dargestellte RNA (b)) müßte dabei durch die lineare mRNA ersetzt werden. Die DNA-Bereiche a) bzw. c) entspräche damit der ausschleifenden Intronsequenz. Nach den entsprechenden Schnitten (Pfeile) und nach Ligation der DNA-Bereiche a) und b), würde eine zirkuläre DNA hervorgehen, welche genau dem Intron entspricht. Die reverse Transkription der gespleißten Intron 1-RNA könnte wegen deren Lariatstruktur nicht zu diesem Plasmid führen.

3.1.5.2.6 INTRONDELETION OHNE DIE BETEILIGUNG VON RNA

Es scheint heute, daß zumindest einige mitochondriale Introns mobile Elemente darstellen (58, 68, 83). In *Sacch.cerevisiae* beinhaltet die Transposition eines Introns des 21S-rRNA Gens (ω -Intron) einen DNA-Doppelstrangsnchnitt an der Insertionsstelle, welcher von einem innerhalb des Introns codierten Proteins katalysiert wird (58, 68). In *Neurospora* wurde ein exzisiertes Intron als stabiles, replikationsfähiges Plasmid gefunden (83). Das α -SEN-Plasmid von *P.anserina* wird ebenfalls durch die Sequenz eines Introns gebildet (vgl. Abschn. 3.1.5.2.3). Die häufigen Deletionen mitochondrialer Introns erinnern zwar an das RNA-Spleißen, die Beteiligung von RNA konnte bisher jedoch nicht nachgewiesen werden.

Da (i) reverse Transkription exzistierender Introns wegen deren Lariatstruktur nicht zu den beobachteten Plasmiden aus Intron DNA führen würde, (ii) die in dieser Arbeit beschriebenen Klasse B-Deletionen Sequenzen betreffen, die Exons entsprechen, mußte eine komplizierte RNA/DNA-Heteroduplex angenommen werden, welche die Ligation der Deletionsenden erlauben könnte (s.o.).

Allen Introndeletionen könnte aber auch ein eigenständiger Mechanismus zu Grunde liegen, welcher sich, ähnlich der Insertion des ω -Introns, vom RNA-Spleißen unterscheidet. Kürzlich gemachte Beobachtungen bekräftigen diese Ansicht (93).

Eine Mutation (G2457) betrifft einen Basenaustausch am 5'Ende des Exon 4 im COB Gen von *Sacch.cerevisiae* (28). Die Mutation führt zu ungenauem Spleißen der RNA (93), und zu einem strikten mit-Phänotyp. Die Rate der exakten Introndeletion wird von der Mutation aber nicht vergleichbar betroffen, weshalb die Autoren die Unabhängigkeit von RNA-Spleißen und DNA-Spleißen vermuten (93).

Die Unabhängigkeit beider Mechanismen würde auch Probleme lösen, welche durch die Forderung von gespleißter RNA in spleißdefekten Mutanten (40) aufgeworfen werden.

3.2 GC-CLUSTER

3.2.1 EINFÜHRUNG

Das mitochondriale Genom der Hefe *Sacch.cerevisiae* zeichnet sich durch einen besonders niedrigen G+C-Gehalt (18%) aus. Frühe Experimente, bei denen das mt Genom mit Mikrokokkus-Nuklease verdaut wurde, konnten zeigen, daß A+T-reiche Spacerregionen, welche über 50% des Genoms bilden, weniger als 5% G+C-Basen besitzen, wogegen etwa 10% des Genoms einen G+C-Gehalt von 60% und mehr aufweisen (94).

Der Verdau mit den Restriktionsenzymen HpaII und HaeIII, welche in den Tetranukleodiden CCGG bzw. GGCC schneiden, ergab, daß die beiden Enzyme an etwa 60 eng umgrenzten Loci des mt Genoms gleichzeitig schneiden (95). Außer diesen HpaII/HaeIII-Clustern wurden 40 weitere HpaII Schnittstellen entdeckt, viele traten innerhalb kurzer Sequenzen mehrfach auf. Beide Enzyme schneiden die A+T-reiche mtDNA sehr viel häufiger als statistisch zu erwarten wäre. Dies gilt nicht für Enzyme, welche andere nur aus G+C bestehende Tetranukleotide erkennen. Aus diesen Beobachtungen konnte geschlossen werden, daß G+C-reiche Cluster ähnliche Sequenzen mit vermutlich palindromer Struktur besitzen.

Sequenzierungen mitochondrialer DNA offenbarten die ersten Primärstrukturen von GC-Clustern (7, 25, 84, 85, 108, 118). Erste systematische Vergleiche von GC-Clustern folgten (43, 107, 102). Es wurde deutlich, daß viele Cluster ähnliche Primärsequenzen besitzen, wobei deren Randsequenzen (5')TAGT und AAGGAG(3') besonders konserviert sind (107). Die bisher umfangreichste Zusammenfassung, welche einen Großteil aller GC-Cluster beinhaltet, wurde kürzlich publiziert (126).

3.2.1.1 VERTEILUNG G+C-REICHER SEQUENZEN IN DER MTDNA

Heute sind etwa 90% der mitochondrialen DNA sequenziert. Die Länge des Genoms variiert in unterschiedlichen Hefestämmen zwischen 75kb und 85kb, es

beherbergt etwa 200 GC-Cluster. Die Positionen von 137 gut erhaltenen GC-Clustern wurden in Abb.1 wiedergegeben.

Mit wenigen Ausnahmen sind alle GC-Cluster Bestandteile nicht-codierender Sequenzen, d.h. vorwiegend in Spacern aber auch in den geschlossenen Leserahmen (CRF) der Introns angesiedelt. Aus der Genomlänge und der Häufigkeit der Cluster errechnet sich das Auftreten eines GC-Cluster etwa alle 400 bp, durch ihre Konzentration in nicht-codierenden Bereichen sogar eine noch höhere lokale Häufigkeit. Die die Cluster umgebende Sequenz ist in der Regel sehr A+T-reich, oder sie stellt ihrerseits ein GC-Cluster dar.

3.2.1.2 DIE TARGETSITE-HYPOTHESE

Durch die in Abschn. 3.1.2 und 3.1.3 erörterte Sequenzierung des wt 777-3A (OXI3-Leaderregion bzw. Intron aI5b) und den Sequenzvergleich mit den wt-Stämmen KL14-4A und D243-2B-R1 konnte gezeigt werden, daß KL14-4A fünf der sieben in 777-3A angetroffen GC-Cluster nicht besitzt (Cluster II, III, IV, VI, VII, in Abb.6 bzw. Abb.8). Abb.18 gibt die polymorphen Genbereiche schematisch wieder.

Anstelle der GC-Cluster war in KL14-4A (GC-Sequenz) das Trinukleotid TAG (oder ein ähnliches Oligonukleotid) zu finden (vgl. Abb.6 bzw. Abb.8), welches zugleich die typische Anfangssequenz der GC-Cluster darstellt.

Diese Beobachtungen führten zur **Targetsite-Hypothese**, wonach das GC-Clustern vorausgehende TAG Trinukleotid nicht die clustereigene Startsequenz repräsentiert, sondern vielmehr die bisher nicht erkannte Zielsequenz ('Targetsite') bei der Insertion der Cluster darstellt. Das AG-Dinukleotid dieser Zielsequenz, welches auch am Ende der GC-Cluster erscheint, könnte während der Insertion der Cluster verdoppelt werden, wodurch GC-Cluster in die Nähe vieler mobiler Elemente rücken, deren Transposition von der obligaten Duplikation von 3-12 bp ihrer Targetsequenz begleitet wird.

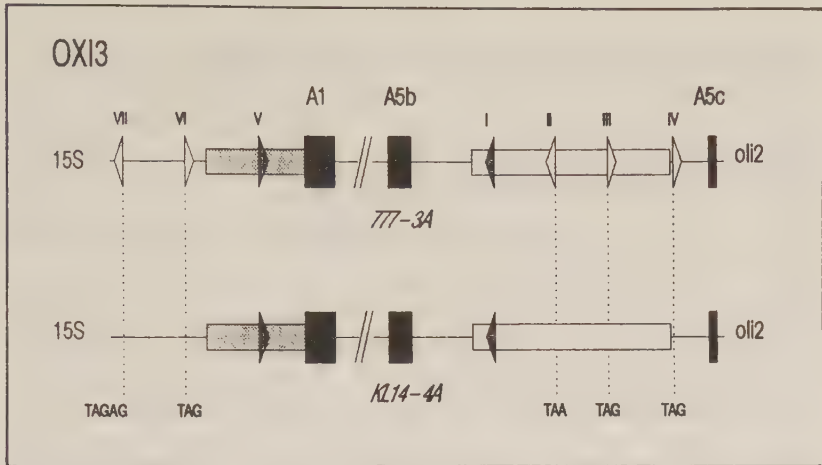


Abb.18 GC-Cluster Polymorphismen im OXI3-Bereich

Polymorphe Bereiche des OXI3-Gen in den wt-Stämmen KL14-4A (obere Zeile) und 777-3A (untere Zeile). Exons sind durch dunkle, der ORF des Introns a15b durch den hellen Balken wiedergegeben. Der schattierte Balken stellt den Leader dar. Dreiecke geben die GC-Cluster und deren Orientierung wieder, wobei helle Dreiecke optionale Cluster bezeichnen. Die Sequenzen der abgebildeten Bereiche wurden in Abb.6 bzw. Abb.8 dargestellt.

Die folgenden Abschnitte befassen sich mit der Überprüfung dieser Hypothese. Dazu wurden alle bisher bekannten GC-Cluster des mt Genoms untersucht, und ein System zu ihrer Einteilung geschaffen. So konnte die Organisation der verschiedenen GC-Cluster Gruppen erkannt werden. Die Bestimmung weiterer Sequenzen, welche sich in verschiedenen Hefestämmen durch optionale GC-Cluster unterscheiden, und im Besonderen die Sequenzanalyse unterschiedlich komplexer GC-Cluster Aggregate trug dazu bei, die wesentlichen Kriterien der GC-Cluster Transposition zu erkennen. Die Targetsite-Hypothese wurde dabei im wesentlichen bestätigt. Um die verschiedenen Gruppen von GC-Clustern zu erfassen, wurden einige Aspekte der Hypothese erweitert.

3.2.2 ENTWICKLUNG EINER GC-CLUSTER-SYSTEMATIK

3.2.2.1 METHODE UND DEFINITION

Zum Studium von Organisation, Sequenzvielfalt und Verwandtschaftsbeziehungen bei GC-Clustern durchsuchte ich alle publizierten und eine Reihe noch unveröffentlichter mit Sequenzen von *Sacch.cerevisiae* nach G+C-reichen Sequenzmotiven. Die Sequenzen dieser G+C-reichen Motive und ihrer flankierenden Nukleotide gab ich in einen Microcomputer ein. Da keine geeignete Software zur Verfügung stand, wurde sie selbst entwickelt (vgl. Abschn. 2.2.2).

Nach einer ersten Analyse der Daten erschienen folgende Beschränkung sinnvoll: G+C-reiche Sequenzen, welche nicht vollständig bestimmt wurden oder weniger als 12 G+C-Nukleotide enthielten, wurden von der weiteren Analyse ausgeschlossen; eine Ausnahme von dieser Regel wird später begründet. Die spezialisierten Cluster A, B und C, welche Bestandteile der sog. Replikationsstart-Sequenzen der mtDNA darstellen (128), werden hier nicht behandelt, obgleich Cluster C mit 14 G+C-Nukleotiden formal den o.g. Kriterien genügen würde. Diese willkürlichen Beschränkungen erschienen nach der Durchsicht der Daten sinnvoll, sie erleichterten das Erkennen der wesentlichen Strukturen. Die Anzahl der zu untersuchenden GC-Cluster verringerte sich damit auf 137.

Die Anfangssequenz (5')TAGT und die Endsequenz AGGAG(3') ist bei der überwiegenden Mehrheit aller GC-Cluster konserviert. Der Mittelteil der Cluster wird durch ein G+C-reiches, nicht perfektes Palindrom gebildet. Beides erlaubt, die Orientierung der Cluster im Genom zu erkennen. Beide möglichen Orientierungen sind etwa gleich häufig. Die Sequenzen aller GC-Cluster wurden in einheitlicher Orientierung ausgerichtet (TAGT->AGGAG), sie stellen damit entweder die Sequenz des transkribierten oder des nicht-transkribierten DNA Stranges dar.

Mit Hilfe des Programms MX (vgl. Abschn. 2.2.2.1) wurde zunächst jedes Cluster mit allen anderen in einer speziellen Homologiematrix verglichen (Daten nicht gezeigt). Die resultierende Sequenzmatrix erlaubte es auf verhältnismäßig einfache Weise, in einer ersten Näherung die Gesamtheit der Cluster distinkten Gruppen zuzuordnen. Zur detaillierten Analyse dieser Gruppen entwickelte ich

das Programm ALN (vgl. Abschn. 2.2.2.2) . Das Programm verwaltet beliebig viele kurze (max. 255 bp) Sequenzabschnitte. Es vermag in Interaktion mit dem Anwender gleiche Bereiche ähnlicher Sequenzen übersichtlich darzustellen und Konsenssequenzen daraus abzuleiten. Die Homologie aller Cluster mit diesen Konsenssequenzen und ihre Organisation bei komplexen Aggregaten aus mehreren GC-Clustern wurde in einem dritten Schritt untersucht.

3.2.2.2 GC-CLUSTER GRUPPEN MIT ÄHNLICHER PRIMÄRSTRUKTUR

Meine Untersuchungen ergaben, daß sich die Gesamtheit aller GC-Cluster, die den oben angeführten Bedingungen genügen, auf nur wenige Gruppen verteilt, diese werden später vorgestellt. Abb.19 zeigt eine Homologiematrix, welche alle 137 Cluster mit den Konsenssequenzen aller Familien vergleicht. Es läßt sich erkennen, daß diese Konsenssequenzen ausreichen, nahezu alle GC-Cluster zu beschreiben. Wegen der Ähnlichkeit einzelner Konsenssequenzen erscheinen einige der Cluster bei oberflächlicher Betrachtung mehreren Gruppen gleichzeitig zuzugehören. Unter Berücksichtigung einzelner, für die jeweiligen Konsenssequenzen typischen Basen (determinierende Positionen s.u.) kann hier jedoch eine Entscheidung getroffen werden. Die Unterbrechungen in den Diagonalen der mit U3 und U4 bezeichneten Spalten entsprechen regelmäßigen Insertionen variierender Länge in den Sequenzen dieser Cluster.

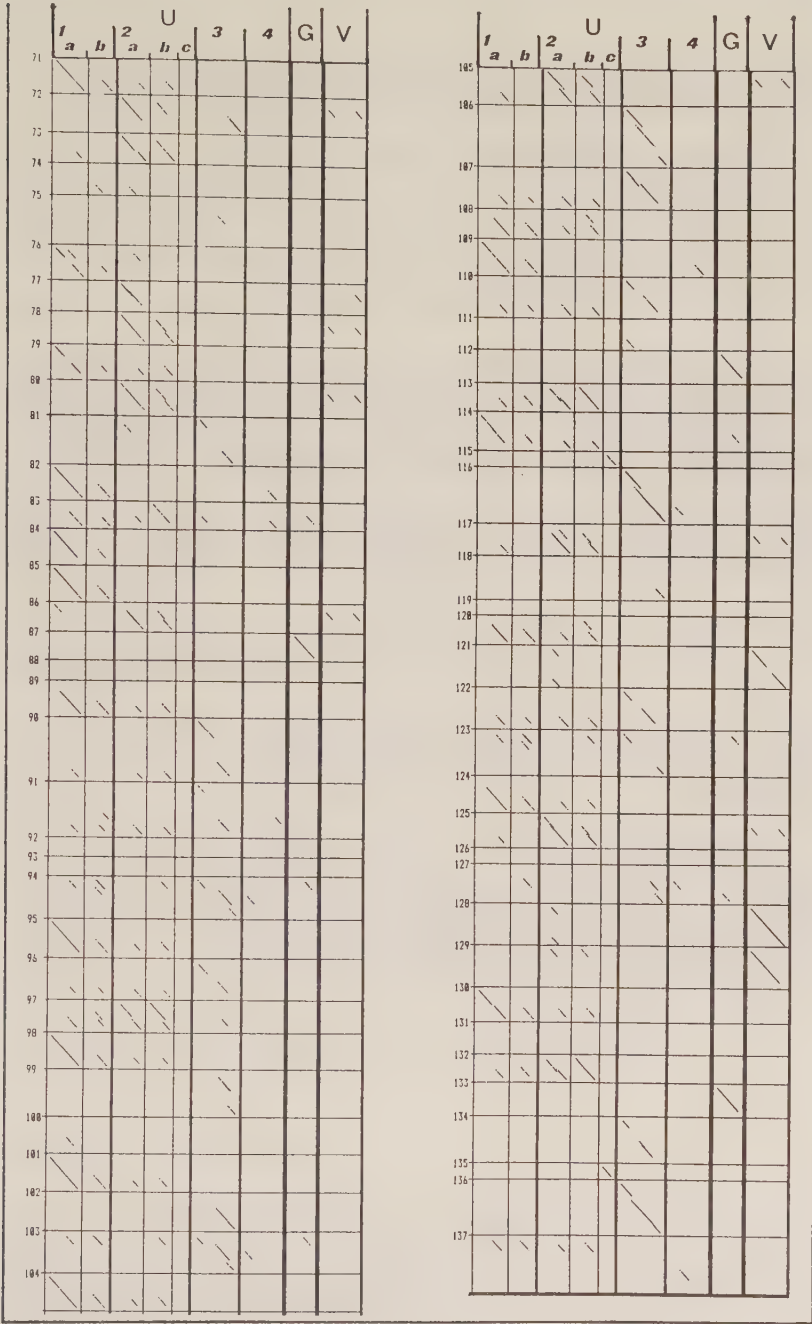


Abb.19 Homologiematrix der GC-Cluster und ihrer Konsenssequenzen

3.2.2.3 SYSTEM DER GC-CLUSTER (EIN VORSCHLAG ZUR NOMENKLATUR)

Auf der Grundlage von Clusterenden, Homologie der Cluster mit den Konsenssequenzen, determinierenden Sequenzpositionen, Homologie der Konsenssequenzen untereinander, typischer Zielsequenzen, in welche die Cluster integrieren können, und weiterer typischer Sequenz- und Strukturmerkmale, schlage ich zur Einteilung der GC-Cluster folgende Nomenklatur vor:

Klasse U (Unit-Cluster) beinhaltet alle GC-Cluster, welche das o.g. Clusterende AGGAG(3') besitzen. Die Klasse umfaßt das Gros aller GC-Cluster und gliedert sich in die Familien U1a, U1b; U2a, U2b, U2c; U3; U4. Mit Ausnahme der Familie U3 beginnen U-Cluster mit der Konsenssequenz TAGT.

Klasse G (Gamma-Cluster) wurden ausnahmslos in Sequenzen gefunden, die mit den Startsequenzen der mtDNA Replikation homolog sind (128, 129).

Klasse V (Var-Cluster) besteht aus einer einzigen sehr homogenen Familie mit nur 4 Mitgliedern. Cluster dieser Klasse wurden im VAR1-Gen beschrieben (56).

Klasse I (Imperfekte Cluster) faßt die restlichen nicht klassifizierbaren Cluster zusammen. Meist handelt es sich dabei um durch Mutation reduzierte Cluster der obigen Klassen.

Im selben Zeitraum, in dem die vorliegende Arbeit entstand, wurden von Zamaroczy & Bernardi systematische Untersuchungen an GC-Clustern vorgenommen. In Abschn. 3.2.3.1 werden die Ergebnisse der beiden von einander unabhängig durchgeführten Arbeiten verglichen.

3.2.3 MERKMALE DER GC-CLUSTER KLASSEN UND FAMILIEN

3.2.3.1 DIE KLASSE U

3.2.3.1.1 NUKLEOTIDSEQUENZEN DER U-CLUSTER

Die GC-Cluster der einzelnen Klasse U-Familien sind in Abb.20 bis Abb.23 gemeinsam mit den Konsenssequenzen der Hauptfamilien wiedergegeben. Die Lokalisation der Cluster und ihre Orientierung im Genom kann im Anhang A nachgeschlagen werden. Wie später gezeigt wird, können die am 5'Anfang isoliert dargestellten Nukleotide nicht als Bestandteile des Clusters aufgefaßt werden. Die einzelnen U-Familien werden im folgenden vorgestellt und in Abschn. 3.2.3.1.6 detailliert diskutiert.

Familie U1a (vgl. Abb.20a)

Mit 31 Mitgliedern ist diese Familie weit verbreitet, sie wurde in der Literatur mehrfach erwähnt und als 'oris-Cluster' (43), 'Klasse A-Cluster' (102) bzw. 'Familie a1' (126) bezeichnet. Die U1a-Cluster bilden die homogenste GC-Cluster-Familie. Acht der Cluster besitzen identische Nukleotidsequenzen, die restlichen Cluster weichen meist nur in einer bis zwei Positionen von ihrer Konsenssequenz ab. Die 9 bp Deletion der Cluster #108 und #120 ist möglicherweise einem HpaII-Fragmentverlust während der Sequenzierungsarbeiten geschuldet. Sieben Cluster sind optional, d.h. in manchen Hefestämmen nicht vorhanden. Ohne Berücksichtigung des (5')TAG-Trinukleotids ist die Konsenssequenz dieser Familie 40 bp lang, ihr G+C-Gehalt beträgt 77,5%.

Familie U1b (vgl. Abb.20b)

Diese Familie enthält nur 4 Mitglieder, sie wurde bereits von Séraphin et al. (102) erwähnt und dort als 'Klasse B' bezeichnet. Eines der U1b-Cluster ist optional. Ohne Berücksichtigung des (5')TAG-Trinukleotids ist die Konsenssequenz dieser Familie 29 bp lang, ihr G+C-Gehalt beträgt 65,5%. Die Bezeichnung U1b wurde

a) U1a	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#13	TAG	TTCCGGGGCCCGGCCACGGAAGCCGGA-CCCCGAAAGGA-							
#17 (*)	TAG	TTCCGGaGCCCGGCCACGGGAGCCGGAACCCCG-AAAGGA							
#25	TAG	TTCCGGGGCCCGGCCACGGGA-CCGGAACCCCGAAAGGAG							
#26	TAG	TTCTGGGGCCCGGCCACGGGAGCCGGAACCCCGtAAGGAG							
#33	TAAAG	TTCCGGGGCCCGGctACGGGAGCCGGAACCCCGAAAGGAG							
#39	TAG	TTCCGGGgttCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#40	TAAg	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#41 (*)	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#42 (*)	GAGGAG	TTCCGGGGCCCGGCCACGGGAACCCGGAACCCCGAAAGGAG							
#48	TAG	TTCCGGGGCCCGGCCACGGgtGCCGGAACCCCGAAAGGAG							
#53 *	TAG	TTCCGGGGCCCGGCCAC-GGAGCC-GAACCC-GAAAGGAG							
#55	GAGGAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#65 *	GAGGAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#66	TAG	TTCCGG-cCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#71 *	TAA	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#76	TAG	TTtCGGGGgCCG ¹ CCA-GGGA-CCGGAACCCCaAAAGGAG							
#79 *	TAG	TTCCGGG-CCCG ² CCA-GGGA-CCGGAACCCCGAA ³ GGAG							
#82	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGcAAGGAG							
#84	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGacAAGGAG							
#85	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGgAAGGAG							
#89	TAG	TTCCGgaG ⁴ CCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#95	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#98	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#101	TAG	TTCCGGGGCCCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#104	TAG	TTCCGGGGCCCGGCcAaCGGGAGCCGGAACCCCGAAAGGAG							
#108	TAG	TaCCGG-----cCGGGAGCCGGAACCCCGAAAGGAG							
#109	TAG	TTCCGGGGCCCGgtCACGGaAGCCGGAACCCCGcAAGGAG							
#114	TAGt	TTCCGGGcCCCGGCCA ⁵ GGGAaCCGGAACCCCGaAGGAG							
#120	TAG	TTCCGG-----cCGGGAGCCGGAACCCCGAAAGGAG							
#124	TAG	TTCCGGGGgttCGGCCACGGGAGCCGGAACCCCGAAAGGAG							
#130	TAG	-TCCGGGGCCCGGCCACGG-A-CCGGAACCCCGAAAGGAG							
b) U1b									
#14 *	T-G	TTCC-----cgcgaaagcg-----GGAACCCCAatAAGGAG							
#27	cAG	TTCC-----cgcttcgcg-----GGAACCCCGtAAGGAG							
#28	TAG	TTCC-----cgcttcgcg-----GGAACCCCGtAAGGAG							
#29	TAG	TTCC-----cgcttcgcg-----GGAACCCCGttAAGGAG							
1	GA	2	GT	3	GA	4	TT	5	CC

Abb.20 U1-Cluster

Von der Konsenssequenz (obere Zeile) abweichende Nukleotide sind in Kleinbuchstaben dargestellt. Optionale Cluster sind mit '*' markiert. Mit '(*)' markierte Cluster sind in Abb.25 nicht berücksichtigt, da die Daten der flankierenden Sequenz fehlen. Vgl. auch Anhang A (siehe Text).

gewählt, um die enge Verwandtschaft der Familie zur Familie U1a zu berücksichtigen. Die Sequenz (5')TTCC am Anfang der Cluster ist beiden Familien gemein, sie stellt ein determinierendes Merkmal der beiden U1-Familien dar. Am 3'Ende der Cluster beider U1-Familien sind 15 von 16 bp identisch, während der Mittelteil spezifisch abweicht.

Familie U2a (vgl. Abb.21a)

Diese Familie enthält 24 Mitglieder; auch sie wurde in der Literatur bereits erwähnt und mit 'oris related' (43) bzw. 'Familie a2' (126) bezeichnet. Fünf U2a-Cluster sind optional. Ohne das (5')TAG Trinukleotid ist die Konsenssequenz der U2a-Cluster 35 bp lang, ihr G+C-Gehalt beträgt 85,7%. Die Familie ist weit weniger homogen als die beiden erstgenannten. Obgleich die einzelnen Cluster nur wenig von ihrer Konsenssequenz abweichen, befinden sich offensichtlich keine zwei völlig identischen Cluster innerhalb der Familie. Die häufigsten Sequenzabweichungen betreffen eine Serie von drei C (Pos.6-8) und eine Serie von fünf C (Pos.10-14). Eigene Sequenzierungsarbeiten ergaben, daß die Interpretation der Sequenzgele wegen einer ausgeprägten Kompression der Banden an genau diesen beiden Stellen besonders problematisch ist. Einige der Cluster konnten nur durch den Ersatz von dGTP durch dTTP korrekt sequenziert werden. Es kann daher vermutet werden, daß nicht alle Sequenzen der Cluster fehlerfrei bestimmt wurden. Doch auch andere Bereiche zeigen Abweichungen von ihrer Konsenssequenz. In vier der Cluster sind bis zu 8 bp inseriert, in sechs Fällen ist mehr als eine Base deletiert. Diese Abweichungen könnten ein höheres phylogenetisches Alter der Familie widerspiegeln. Die Primärsequenz der U2a-Cluster ähnelt jenen der U1-Familien. Als abweichende Indikatorpositionen können u.a. betrachtet werden: Das 5'Ende TCC (U1=TTCC) und das 3'Ende GAAGGAG (U1a=GAAAGGAG), also ein fehlendes T (3'Ende) und ein fehlendes A (5'Ende), ferner ein fehlendes A in der Sequenz CGGA 12 nt vor dem Clusterende (U1a=CGGAA).

Familie U2b (vgl. Abb.21b)

Sechs Cluster zählen zu dieser Familie, wobei zwei optional sind. Vier Cluster sind in sog. Replikationsstart-Sequenzen inseriert (ori4, ori6, ori1*) diese Cluster

a)	U2a	TAG	TCCGGCCCCGCCCCCGCGGGGCGGACCCCGAAGGAG
#6		TAG	TCCGGCC-GCCct-GCGGGGCGGACCCCGAAGGAN
#7		TAG	TCCGGCC-GCCct-GCGGGGCGGACCCCG-gGtcG
#8		TAG	TCCGGCCCCGCCCC-GC--GGCG---CCCGAAGGAG
#9		GAG	TCCGGCCCCGCCCCCGCGGGGCGGACCCCGAAGGAG
#16		TAG	TCCGctCCcctCCCCG ¹ GGGCGGACCCCGAAGGAG
#22		TAG	TCCGGC----CCCCGCGGGGCGGACCCCGAAGGAG
#30		TAG	TCCGGCC-GCCCC-GtGGGGCGGACCCCaAAGGAG
#38		TAG	TCCGGCCCCGCCCC-GC-GGGCGGACCCCGcAGGAG
#52		TAG	TCCGaCggcCCCCGCGGGGCGGACCCCGAAGGAG
#57		TAG	TCCGGCC-GCCCCCGCGGGGCGtACCCCGAAGGAG
#60	*	TAG	-gattCCCaCCCC-aGtGG-GGACCCCGAAGGAG
#63	*	TAG	TgCGGCCCGCCCCCGCGGGGCGGACCCCGAAGGAG
#68	*	TAG	TCCGGCCCCGCCCCCGCGGG ² CGGACCCCGAAGGAG
#72		TAG	TCCGGCCCCGCCCCCGCGGGGCGGACCC ³ AAGGAG
#73		TAG	TCCGGCCCCGCCCCCGC--GGCG-ACCCCGAAGGAG
#77	*	TAG	TCCGGCCCCGCC--GCGGGGCGGA ⁴ CgCCGgAGGAG
#78	*	TAG	TCCGGCCCCGCCCCCGCGGGGCGGACCCCaAAGGAG
#80		TAG	TCCGgtCCGCCCCCGCGGGGCGGACCCCaAAGGAG
#86		TAG	TCCGGC--GCCCC-GCGGGGCGGACCCCaAAGGAG
#105		TAG	TCCGGCCCCGCCCC-GCGGG ⁴ CGGACCCCGAAGGAG
#105'		TAG	TCCGGCCCCGCCCC-GCGGG ⁴ CGGACCCCGgAGGAG
#117		TAG	TCCGGC-gGCCCCCGCGGGGCGGACCCCGAAGGAG
#125		TAG	TCCGGCCaGCCCC ⁵ GCGGGGCGGACCCCGAAGGAG
b)	U2b		
#45'	*	TAAAG	TCC---CCGCCCCgGCGGGG---ACCCCGAAGGAG
#45		GGG	TCC---CCGCCcGgGCGGGG---ACCCCGAAGGAG
#70	*	TAGAG	TCC---CCGCCCCgGCGGGG---ACCCCGAAGGAG
#83		TAG	TCC---CgGCCCCaGCGGGG---ACCC ⁶ AAGGAG
#97		TAAAG	TCC---CCGCCCCgGCGGGG---ACCCG ⁷ AAGGAG
#113		TAAAG	TCC---CCGCCCCgGCGGGG---ACCCCGAAGGAG
#132		TAAAG	TCC---CCGCCCCgGCGGGG---ACCCCGAAGGAG
c)	U2c		
#5	GAAGGAN	TCCGaCC-----	GAAGGAa
#10	TAAG	TCCc-----	aAAGGAG
#18	TAG	TTCCG-----	GAAGGAG
#20	GAAGGAG	TCCGGCC-----	GAAGGAG
#50	GAAGGAG	TCCGGCC-----	GAAGGAG
#56	GAAGGAG	TCCGgtC-----	GAAGGAG
#56'	TAG	TCCGGCC-----	GAAGGAG
#115	GAAGGAG	TCCGgtC-----	GAAGGAG
#119	TAG	TTCCtt-----	GAAGGAG
#126	GAAGGAG	TCCGaCt-----	GAAGGAG
#135	TAA	TCCGGCC-----	GAAGGAG

1 AGGGGAGG 2 GGGGG 3 TTTTAAAG 4 GGGGGGG
5 CCC 6 GC 7 AACCCCG

Abb.21 U2-Cluster

Zur Legende siehe Abb.20.

wurden von verschiedenen Autoren als β -Element bezeichnet (126, 43). Ihre Klassifizierung als U2b-Cluster berücksichtigt ihre weitgehende Sequenzhomologie mit den Clustern der U2a-Familie. Aus diesen können sie formal durch die Deletion der beiden Trinukleotide GGC bzw. GCC und einer Transition C->G abgeleitet werden. Ohne das einleitende TAG-Motiv ist die Konsenssequenz der Familie 29 bp lang, ihr G+C-Gehalt beträgt 82,8%.

Familie U2c (vgl. Abb.21c)

Die Konsenssequenz der 10 Cluster dieser Familie ist nur 14 bp lang, ihr G+C-Gehalt beträgt 71,4%. Damit erfüllen U2c-Cluster als einzige Ausnahme nicht das o.g. Kriterium (12 G+C-Basen). Folgende Merkmale weisen sie jedoch als eigenständige GC-Cluster-Unterfamilie aus. (i) Fünf Cluster sind an der Formation von Clusteraggregaten beteiligt (s.u.). (ii) Die Clusterenden, welche die typischsten Strukturelemente von U-Clustern darstellen, sind völlig erhalten. (iii) Drei der Cluster besitzen identische Nukleotidsequenzen. Diese Cluster könnten durch Deletion des zentralen G+C-Palindroms von U2a-Clustern entstanden sein. Einige der Cluster könnten sich auch aus U1a-Clustern ableiten. Bei der Analyse der mtDNA-Sequenzen wurden noch kürzere Elemente gefunden (z.B. TAGGAG, Daten nicht gezeigt), diese werden hier jedoch nicht als GC-Cluster behandelt.

Familie U3 (vgl. Abb.22)

Zu dieser verbreiteten Familie zählen 34 Cluster, sechs sind optional. Auf einige dieser Cluster wurde in der Literatur bereits hingewiesen (34) und bei (126) als 'a3-Familie' bezeichnet. An zwei spezifischen Positionen werden viele der Cluster von typischen A+T reichen Sequenzmotiven unterbrochen, diese Motive sind oft mehrfach wiederholt. Das häufigste Monomer an Pos. 15 stellt die Sequenz CTATACT dar. Das zweite Insert (TTTTAAA od. ähnlich) liegt unmittelbar vor dem Clusterende (GAAGGAG). Im Gegensatz zu den übrigen U-Clustern geht den U3-Clustern nicht das Trinukleotid TAG voraus, sondern das typische Endmotiv (AAGGAG) der U-Cluster-Familien. Nur in einem Fall (#81) wird ein U3-Cluster durch das TAG-Trinukleotid eingeleitet. Ohne Berücksichtigung dieser einleitenden Sequenzen und ohne die o.g. A+T-reichen Inserts ist die Konsenssequenz der U3-Cluster 43 bp lang, ihr G+C-Gehalt beträgt 67,4%.

a)	U3	TCAGGACCCCTCC .ctatact..... ATGGAGGGGGACCGAACCCC .tttaaa.....CAAGGAG
#1	*	AAGGAG TCAGGTGCCCCTCC .CTTAT (ATTATA) 3ATCGTACTT.ATGGGGGGGGAACCGAACCCC.TT (TTTAA) 2. TTAA. .AAGGAG
#11		CGAAGG TCAGAGGGCTCTCTCC.ATACATCA.....ATGG-AGGG--ACCGAACCCC.....CAAGGAG
#15		AAGGAG TCGGGACCCC---ATGGGA---ACCGCATCC.TTTTATTC (TTAA) 2. GAAGGAG
#21		AAGGAG TCAGGACCCCTCC.. (CTATACT) 2. A.....ATGGAGGGGACCGAACCCC. TTTTAAA.....GAAGGAG
#24		GAGGGT TGAGGACCCCTCC..CTATACTA.....ACGGAGGGGACCGAACCCC.....GAAGGAG
#44	*	AAGGAG TCAGGACCCCTCC. (CTATACT) 2. A.....ATGGAAGGGGACCGAACCCC.....GAAGGAG
#49		AAGGAG TCAGGACCCC-C-..GTATACTT.....ACGG--GGGACCGAACCCC.....GAAGGAG
#51		AAGGAG TCAGGACCCCTCC- (CTATACT) 2. A.....ATGGAGGGGACCGAACCCC. TTTTAAA.....GAAGGAG
#58	*	AAGGAG TCAGGACCCCTCC. CTATACTA.....ATGGAGGGGACCGAACCCC.....GAAGGAG
#61	*	AAGGA- TCAGGACCCC---GTGGGGACCGAACCCC. TTTTAA.....GAAGGAG
#62		AAGGAG TCAGGACC--TG-...CTA.....ACGGAGGGGACCGAACCCC.....GAAGGA-
#67	*	AAGGAG TCAGGACCCCTCC..TA.....ACGGAGGGGACCG-ACCCC.TTTTATCT (TTAA) 2. GAAGGAG
#90		AAGGAG TCAGGACCCC-CC..G (TATACTA) 4. TATACTT.....AC--GAGGGGACCGAATCCC.....GAAGGAG
#91		AAGGAG TCAGGACTCTT--TATATT (CGTATC) 2 CGTACCCGT. AAGGGAG--ACCGAACCCC.....GAAGGAG
#102		AAGGAG TCATAGATCCT--TT.....-----GGGGACCGAACCCC. TATTTAA.....GAAGGAG
#103		AAGGAG TCGGGACCCC--G.....TGGA---CCGAACCCC. TTTTATTTTAAA...GAAGAAG
#106	*	AAGGAG TCAGGACCCCTCC.. (CTATACT) 2. A.....ACGGAGGGGACCGTATCCC. (TTTTTA) 2. A.....GAAGGAG
#107		AAGGAG TCAGGACCCCTCC. CTA.....ATGGAGGGGACCGAACCCC.....GAAGGAG
#107'		GAGGAG TCAGGACCCCTCC. CTATCTA.....ACGGAGGGGACCGAACCCC.....GAAGGAG
#116		AAGGAG TCAGGACCCCTCC.. (CTATACT) 2. A.....ATGGAGGGGACCGAACCCC. TTTTAAA.....GAAGGAG
#122		AAGGAG TCAGGACCCCTCC. GTT.....A-GGGAGGGGACCGAACCCC.....GAAGGAG
#122'		AAGGAG TCAGGATCCCTCC..GTT.....A-GGGAGGGGACCGAACCCC.....GAAGGAG
#123		AAGGAG TCGGGACCCC---GTGGGA---ACCGCATCC. (TTTTTA) 2. ATTA. .GAAGGAG
#127		AAGGAG TCGGACCCC---ATGGGA-----ACCC. TTTTATTC. TTAA. .GAAGG-G
#136		AAGGAG TCAGGACCCCTCC...C. ATACT. CTATACTA.....ATGGAGGGGACCGAACCCC. TATTTAA.....GAAGGAG

b)	ACGTAAGGAG TT--GGACCCCCCC..CCCG... (TATACT) 2. TA.....-----AGGGGA---AACCCC.CTTTATTC. (TTAA) 2. GAAGGAG
#4	TTTGAATGAG TGAGG-----TGAGGACCGAACCCC.TT (TTTAA) 3. TTAA..GAAGAAG
#31	TATTAAGGAG TGAGGACCCCTCC.CTA.....ACGGAGGGGGACCGAACCCC.....GACCCC.....AAAGGAG
#34	TATTAAGGAG TGAGGACCCCTCC.CTA.....ACGGAGGGGGACCGAACCCC.....AAAGGAG
#35	TATTAAGGAG TGAGGACCCCTCC.CTA.....ACGGAGGGGGACCGAACCCC.....AAAGGAG
#36	TAGAAAGGAG TGAGG-----CC...CTA.....ACGGAGGGGGACCGAACCCC.....GAAGGAG
#81	ATATAAATAG TGAGGGA--CCCTCC..CCCG... (TATACT) 2. TACGG...--GGGAGG--ACCAACCCC.T.....TAG
#94	TTAGTAGGAG TGCGGGACCCCG--.....TGGGA-----ACCGAACCCC.TT (TTTTA) 2. AA.....GAAGAAG
#96	AAGAAAGGAG TGAAGGGACCCCCC..CTT.....ACGGG--GGGAAACCGAACCCC.....GCAGGAG
#110	AAAAAAGGAG TGAGGACCCCTCC..CGT.....TGGGTGGGGACCGAACCCC.....GAAGGAG
#118	TCAGAAGGAGTGGGATACCCAT....GAGG.....ATG--AGGAAACCGAATC--.TTTTAAT.....-AAGGAG
#134	TCAAAAGGAG T-AGGGACCCCTCC..CGTAAATAT.....ACGGAGGGGGACCGAACCCC.....-AAGGAG

Abb.22 U3-Cluster

a) U3-Cluster, welche Bestandteil komplexer Cluster-Aggregate darstellen; b) einzeln angetroffene U3-Cluster. "a" bezeichnet optionale Cluster. Die Bereiche um die beiden A+T-reichen Inserts sind mit Punkten gekennzeichnet. Vgl. auch Anhang A (siehe Text).

Abweichungen der einzelnen Cluster von dieser Konsenssequenz sind häufiger als in allen anderen GC-Cluster-Familien, wobei die meisten Sequenzabweichungen in der Umgebung des ersten Inserts beobachtet werden.

Familie U4 (vgl. Abb.23)

Diese homogene Familie enthält nur vier Cluster, welche sich weitgehend von den restlichen U-Clustern unterscheiden, sie wurde meines Wissens in der Literatur bisher nicht erwähnt. Auf Grund ihrer typischen Clusterenden (5')TAGT u. AGGAG(3')) sind sie der U-Klasse zuzuordnen. Auch U4-Cluster besitzen ein A+T-reiches Insert, wobei hier eine TTTA-Sequenz zwei- bis siebenmal wiederholt ist. Das Insert wird durch das inverse Motiv TAAA beendet. Ohne Insert und ohne das einleitende TAG-Trinukleotid beinhaltet die Konsenssequenz der Familie 41 bp, ihr G+C-Gehalt beträgt 58,5%. Alle U4-Cluster sind mit weiteren Clustern zu Tandemaggregaten verbunden. Zwei dieser Cluster sind optional, wobei in den clusterfreien Allelen stets das komplette Tandem, welches aus einem U4- und einem U1-Cluster besteht fehlt (s.u.).

U4	TAG TGAACACC (TTTA) _n ATTATAAA GGTGTGAACCAATCCC (GCAAG) ₂ GGAGGAG
#3 (#3')	TAG TGAACACC (TTTA) ₇ ----TAAA GGTGTGAACCAATCCC GCAAG GGAGGAG
#43 (*)	TAG TGAACACC (TTTA) ₂ ATTATAAA GGTGTGAACC----CC (GCAAG) ₂ GGAGGAG
#54	TAG TGAACACC (TTTA) ₂ ATTATAAA GGTGTGAACCAATCCC (GCAAG) ₂ GGAGGAG
#64 *	TAG TGAACACC (TTTA) ₂ -----AA GGTGTGAACCAATCCC (GCAAG) ₂ GGAGGAG

Abb.23 U4-Cluster

Zur Legende vgl. Abb.22

3.2.3.1.2 TARGETSEQUENZEN OPTIONALER U-CLUSTER IN A+T-REICHEN SEQUENZEN

Die Entdeckung fünf optionaler Cluster führte zur Targetsite-Hypothese (vgl. Abschn. 3.2.1.2). Zur Überprüfung dieser Hypothese wurden weitere Sequenzen bestimmt, welche sich nur durch ihre GC-Cluster unterschieden (vgl. Abschn. 3.2.3.1.5) und alle in der Literatur verfügbaren Sequenzen aus unterschiedlichen Hefestämmen nach weiteren GC-Cluster-Dimorphismen untersucht. Zusätzlich zu den o.g. fünf optionalen (Klasse U) GC-Clustern wurden weitere 11 Sequenzdimorphismen gefunden, in welchen ein U-Cluster (8 Fälle) oder zwei U-Cluster in Tandemformation (3 Fälle) in A+T-reiche Sequenzen inseriert waren. Bei drei optionalen U-Clustern war zumindest eine der an das Cluster anschließenden Sequenzen nicht bestimmt worden, ihre Insertionsstellen können daher nicht eindeutig festgestellt werden. Abb.24a,b vergleicht die verbleibenden 13 GC-Sequenzen mit den entsprechenden GC⁺-Sequenzen. (Weitere, komplexere GC-Cluster-Dimorphismen sind in Abb.28 dargestellt; sie werden später diskutiert).

Ein Vergleich der zueinander homologen Replikationsstart-Sequenzen (ori1 bis 8) zeigt außerdem, daß diese Sequenzen sich durch das Vorhandensein zweier U2b-Cluster unterscheiden, welche an spezifischen Positionen enthalten sein können (35, 128). Auch diese beiden Cluster können damit als optional betrachtet werden. Abb.24c zeigt exemplarisch für die beiden Cluster den relevanten Abschnitt zweier Replikationsstart-Sequenzen mit und ohne U2b-Cluster.

Die Ergebnisse dieser Sequenzvergleiche bestätigen die Targetsite-Hypothese. Alle 16 GC-Sequenzen enthalten das TAG oder ein verwandtes Oligonukleotid an der Insertionsstelle der Cluster. Der (5')TAG-Konsens der Cluster ist manchmal erweitert zu TAAG, TAAAG, oder TAGAG; entsprechend findet sich dieses erweiterte Motiv auch in den GC-Sequenzen wieder. Bemerkenswert ist, daß die Orientierung der Cluster stets durch die Orientierung der Targetsequenz determiniert wird.

Die flankierenden Sequenzen sind im allgemeinen sehr A+T-reich, besitzen jedoch keinen darüber hinausgehenden Konsens. Einzelne Sequenzabweichungen wurden festgestellt. Dies war zu erwarten, da Spacerregionen, welche hier vorwiegend betroffen sind, in unterschiedlichen Hefestämmen weniger konserviert sind als codierende Sequenzen. Es scheint jedoch erwähnenswert, daß kleine Abweichungen des TAG (TAAG, TAAAG, TAGAG) Konsens in den GC-

a) optionale U-Cluster	
#79 ⁻ #79 (IV)	GTGTTAAAAATATATCTTTAG-----AATAAAAAATGAATTAAATTA GTGTTAAAAATATATCTTTAGTTCGGGGCCC- (1a) -CCCCGAAGAGGAGATAAAAAATGAATTAAATTA
#71 ⁻ #71 (VI)	A-TAAT-TA--TTA-TA-TAG-----TA-TGA-TA-TATAAA-TCAA AtTAATtTatcTTATatTAaTTCGGGGCCC- (1a) -GGCCCCAAAGGAGTAaTGaATaATATAAAaTCAA
#53 ⁻ #53	AAAATATTTTATTTTATTTat-----AAATATTAATATTTTATAATA AAAATATTTTATTTTATTTAGTTCGGGGCCC- (1a) -AACCCGAAAGGAGAAATATTAATATTTTATAATA
#137 ⁻ #137	CTTTCcATTtTTTA--TTTAAG-----AAAGTATTATTGTTTTCTTT CTTTCcATTtTTATTTTAAAGTTCGGGGCC- (1a/4) -CCGAGAGAGGAGAAAGTATTATTGTTTTCTTT
#14 ⁻ #14	ATTTATAATAATATATATAG-----TTTTATTATTAATTATATTTA ATTTATAATAATATATAT-GTTCcCGCGAA- (1b) -ACCCATAAGGAGTTTTATTATTAATTATATTTA
#78 ⁻ #78 (III)	AAAAGATAGTTATATTTTAG-----ATATTATTAATAAATATTTTA AAAAGATAGTTATATTTTAGTTCGGGGCCC- (2a) -GACCCCAAGGAGATATTATTAATAAATATTTTA
#77 ⁻ #77 (II)	CAGCTGTTAATTCCTTTATTAa-----AATTATATTTAGCTGTAATAT CAGCTGTTAATTCCTTTATTTAGTTCGGGGCCG- (2a) -GACGCGGAGGAGAAATTATATTAGCTGTAATAT
#63 ⁻ #63	AATAATATATTAATAaTAAAG-----TATTATAAAATTAATAATAT AATAATATATTAATAaT-AGTCCGGCCCGC- (2a) -GACCCCAAGGAGTATTATAAAATTAATAATAT
#60 ⁻ #60	AAATTATATAAATATATAgAG-----ATATTATTTTAAATAgAgGT AAATTATATAAATATAT-AGGATTCcCACC- (2a) -GACCCCAAGGAG-TATTATTTTAAATAgAgGT
#70 ⁻ #70 (VII)	AAAATAATATTTAATATAGAG-----TATTATATTTTATATAATTAA AAAATAATATTTAATATAGAGTGTCCCGCCC- (2b) -GACCCCAAGGAGTATTATATTTTATATAATTAA
#45' ⁻ #45'	ATTATATAATAATA-ATAAG-----TATAAACGATATAATTAATTA ATTATATAATAAAATATAAAGTCCCGCCCC- (2b) -GACCCCAAGGAGTATAAACGATATAATTAATTA
b) optionale U-Cluster-Tandem	
#68 ⁻ , 67 ⁻ #68, 67	TATTAACGAAATAATAAAtTG-----TATAATATAAATTATAAATAA TATTAACGAAATAATAAATAGTCCGGCCCG- (2a) (3) -TAATTAAGGAGTATAATATAAATTATAAATAA
#64 ⁻ , 65 ⁻ #64, 65	AAGTAAAAATAATATTAATAG-----AAATTATACAATAATAATAAT AAGTAAAAATAATATTAATAGTGAACACCT- (4) (1a) -CCCGAAAGGAGAAATTATACAATAATAATAAT
c) Replikationsstart-Sequenzen mit und ohne U-Cluster	
ORI5 (GC ⁻)	ATTATATAATAAAATATAAAG-----TATTAACAATATAATATATTA
ORI6 (GC ⁺)	ATTATATAATAAAATATAAAGTCCCGCCCC- (2b) -GACCCCAAGGAGTATTAACAATATAATATATTA

Abb.24 Targetsequenzen optionale U-Cluster in A+T reichen Bereichen

Nur Endsequenzen und Familienzugehörigkeit der GC-Cluster sind gezeigt. GC-Cluster freie Allele (GC⁻-Sequenzen) sind mit der Nummer des Cluster und negativem Exponent gekennzeichnet. Die Nukleotidsequenzen an den Insertionsstellen der Cluster sind hervorgehoben. Vgl. auch Anhang A (siehe Text).

Sequenzen (4 Fälle) häufiger zu beobachten sind als in den GC⁺-Sequenzen (2 Fälle). Die auffällige Konservierung des TAG-Trinukleotids innerhalb der GC-Cluster kann auch bei der Betrachtung der Gesamtheit der GC-Cluster (optionale und nicht optionale U1-, U2-, U4-Cluster) erkannt werden (vgl. Abb.20, 21, 24).

Insertionen von Clustern aus den Familien U2c und U3 konnten hier nicht beobachtet werden. Die Zielsequenzen aller GC-Cluster-Familien werden später gesondert diskutiert (Abschn. 3.2.3.1.6).

3.2.3.1.3 ORGANISATION DER U-CLUSTER IN KOMPLEXEN TANDEMAGGREGATEN

22 G+C-reiche Sequenzen bestehen aus Tandemaggregaten von zwei bis vier GC-Clustern. Ihre einzelnen Elemente entsprechen typischen Clustern aus den verschiedenen U-Familien. Alle GC-Clusteraggregate zeigen ein einheitliches Organisationsprinzip (vgl. Abb.25).

Folgende drei Organisationsmerkmale zeichnen GC-Clusteraggregate aus:

- (i) Nur dem jeweils ersten Cluster eines Aggregates geht das TAG-Trinukleotid voraus, während alle folgenden Cluster unmittelbar an das AGGAG-Ende des vorausgehenden Clusters anschließen.
- (ii) Die Mitglieder der einzelnen Familien nehmen typische Positionen innerhalb eines Aggregates ein. U2a-Cluster und U4-Cluster stehen immer am Anfang eines Aggregates, während U1a-, U2c- und U3-Cluster weiter hinten liegende Positionen einnehmen. U1a-Cluster kommen dabei nur hinter U4-Clustern zu liegen, wobei alle U4-Cluster des Genoms von U1a-Clustern oder einem mit U1a verwandten imperfekten (Klasse I) Cluster gefolgt werden.
- (iii) Innerhalb eines Aggregates besitzen alle Cluster einheitliche Orientierung. Dies erinnert an die bei optionalen Klasse U-Clustern gemachten Beobachtung, wobei die Orientierung der Targetsequenzen die Orientierung der GC-Cluster bestimmte.

#22-19	ATAAATATAG (TCCGGCCCC-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) /ATCGAATC-14-CTAGAGGAG) TAATTATTTA
#52-49	GTTAAATATAG (TCCGAGCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) AATATT
#117-115	TATTTAATATAG (TCCGGGGCC-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) AATATT
#105-107	ATAATATATAG (TCCGGGGCC-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#59-58	ATAATATATAG (TCCGGGGCC-12a/3-AAAGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#16-15	ATAAATATAG (TCCGGTCCC-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#68-67	TAAATATATAG (TCCGGGGCCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#125-127	TATATATATAG (TCCGGGGCCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#6-5	TAAATATATAG (TCCGGGGCCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#57-56	TGAATATATAG (TCCGGGGCCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#135-136	ATAATATATAG (TCCGGGGCCG-2a-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTAATATAT
#102-103	TAGGAGGAG (TGATAGATCC-3-TAAGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTTTATTTT
#62-61	TAGGAGGAG (TGAGGGACCC-3-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTTTATTTT
#91-90	TAGGAGGAG (TGAGGGACCC-3-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTTTATTTT
#122-123	AAGAAGGAG (TGAGGGATCC-3-CCCGAAGGAG) (TGAGGGACCC-3-AAAGAAGGAG) (TCCG-2c-GAAGGAG) (TGAGGGACCC-3-CCCGAAGGAG) TTTTATTTT
#64-65	ATAATATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#54-55	ATAATATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#43-42	ATAATATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#3-1	TATTTATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#23-24	AAAAATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#12-11	AAAAATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA
#46-44	TAAATATATAG (TGAACACCTT-4-AGGGAGGAG) (TTCGGGGCC-1a-CCGAAAGGAG) AAATATACA

Abb.25 GC-Cluster-Aggregate

Nur Endsequenzen und Familienzugehörigkeit der GC-Cluster sind gezeigt. Vollständige GC-Cluster sind durch Klammern begrenzt; unvollständige Clusterenden sind mit '/' markiert. Die Anfangssequenz (TAG) und das typische Endmotiv (AAGGAG) der GC-Cluster sind hervorgehoben (siehe Text).

3.2.3.1.4 REKOMBINATIONSORTE DER CLUSTER-AGGREGATION

Die im vorangehenden Abschnitt gezeigten GC-Cluster-Aggregate entstanden offensichtlich durch Addition einzelner Cluster, wobei die Insertion des dazukommenden Clusters entweder in die (5')TAG Anfangssequenz, oder in die AGGAG(3')-Endsequenz des bereits vorhandenen Clusters (Akzeptor) erfolgt sein kann. In Abb.26 sind beide Möglichkeiten dargestellt.

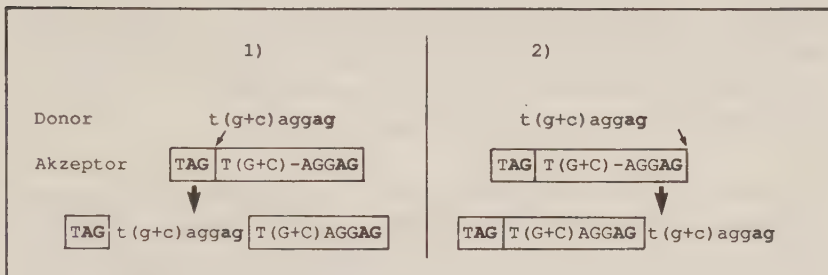


Abb.26 Mögliche Rekombinationsorte der GC-Cluster-Aggregation

Die Abb. zeigt die zwei Möglichkeiten der Integration eines Donor-GC-Clusters (Kleinbuchstaben) nach einem **AG**-Dinukleotid (fett) des Akzeptor-Clusters (Großbuchstaben). Das terminale **ag**-Dinukleotid am 3'Ende des hinzukommenden Donor-Clusters könnte auch am 5'Anfang des Clusters liegen, das Cluster müßte dann vor den **AG**-Dinukleotiden des Akzeptor-Clusters inserieren. Eine dritte Möglichkeit besteht in der Duplikation des Target-**AG**s, das hinzukommende Donor-Cluster müßte dann kein **ag** besitzen (siehe Text).

Beide potentiellen Insertionsstellen werden durch ein **AG**-Dinukleotid markiert, welches während der Insertion eines Clusters möglicherweise verdoppelt wird. Die Targetsequenz (TAG) am 5'Anfang der Cluster ist mit jener Sequenzen identisch, welche von U-Clustern bei der Integration in A+T-reiche Sequenzen benützt wird.

Die Frage, welche der beiden Insertionsstellen (oder beide) von GC-Clustern benützt wird, konnte durch die Analyse innerhalb der Aggregate optionaler Cluster geklärt werden.

3.2.3.1.5 TARGETSEQUENZEN OPTIONALER U-CLUSTER IN GC-CLUSTERAGGREGATEN

Zur Unterscheidung der beiden o.g. möglichen Insertionsstellen wurden Sequenzdimorphismen in den GC-Clusteraggregaten unterschiedlicher Hefestämme gesucht. In meiner Sammlung mitochondrialer Sequenzen konnte ich vier polymorphe Aggregate finden; sie sind in Abb.28 dargestellt (Cluster #1, #62, #44, #123) und werden später diskutiert.

Um weitere Dimorphismen in Clusteraggregaten zu finden, wurde in Zusammenarbeit mit C.Schüller und J.Zimmermann mtDNA von 15 unterschiedlichen Hefestämmen präpariert, und in Southernexperimenten analysiert. Zur Hybridisierung wurden verschiedene mtDNA-Proben verwendet, welche die flankierenden A+T-reichen Sequenzen bekannter Clusteraggregate enthielten.

Bei einigen der Stämme hybridisierten Restriktionsfragmente, welche der aus publizierten Sequenzen voraussagbaren Länge entsprachen, während bei etwa der Hälfte der Stämme kürzere Fragmente auftauchten (Daten nicht gezeigt). Die Größenunterschiede vieler der kürzeren Fragmente (40-70 bp) deuteten auf die Abwesenheit einzelner Cluster in den Aggregaten der kürzeren Stämme hin. Bei einem der kürzeren Stämme (D8-25) wurden die betroffenen Genbereiche (COB Intron 3 bzw. der an die 15S rRNA anschließende Bereich) durch weitere Restriktionsanalysen überprüft. Die Ergebnisse ließen vermuten, daß fehlende GC-Cluster die unterschiedlichen Fragmentlängen bedingten.

Vier der kürzeren Fragmente aus dem wt D8-25 wurden in M13 kloniert und sequenziert. Abb.27a,b gibt die Restriktionskarten der betroffenen Genbereiche und die Strategie ihrer Sequenzierung wieder.

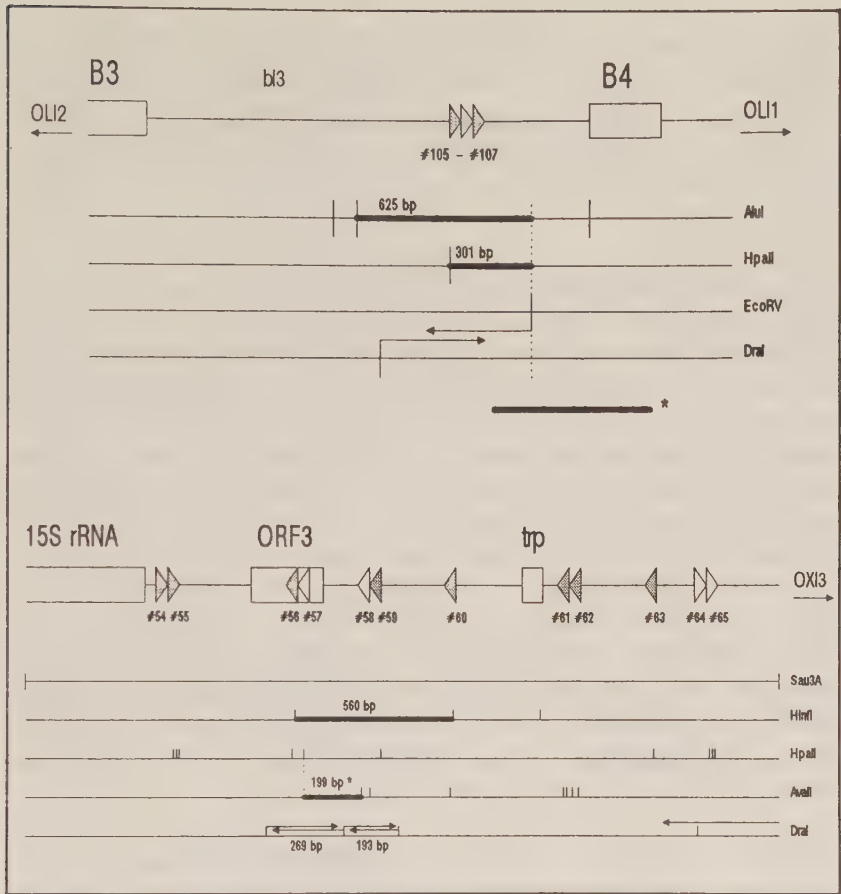


Abb.27 GC-Cluster-Polymorphismen im COB-Intron 3 und nach dem 15S rRNA-Gen :Genomstruktur und Sequenzierungsstrategie

Die langen Formen der beiden Genbereiche sind gezeigt. Helle Balken entsprechen codierenden Sequenzen. Dreiecke bezeichnen GC-Cluster und deren Orientierung. Die hell dargestellten Cluster sind im kurzen Stamm D8-25 nicht vorhanden. Pfeile geben Umfang und Strategie der Sequenzierung von D8-25 wieder. Mit '*' markierte DNA Fragmente wurden als Hybridisierungsproben verwendet. In Southernexperimenten unterschieden sich die durch dicke Linien hervorgehobenen Restriktionsfragmente in langen und kurzen Stämmen. Die Abwesenheit des Clustertandems #64/#65 wurde vermutet, da im kurzen Stamm D8-25 die im Cluster lokalisierte DraI-Restriktionschnittstelle fehlte (siehe Text).

Wie erwartet, entsprachen die Fragmentunterschiede polymorphen GC-Cluster Aggregaten. Drei U3- und ein U2a-Cluster bilden die optionalen Bestandteile der drei längeren Fragmente. Das gemeinsam fehlende Clustertandem (#64/#65) wurde bereits in Abb.24b vorgestellt. Die übrigen polymorphen GC-Clusteraggregate sind in Abb.28 wiedergegeben (Cluster #56/#57, #58/#59, #105/#106/#107).

Aus der Abb.28 wird ersichtlich, daß alle optionalen U3-Cluster erst mit der Sequenz TGAGGG beginnen, wobei sie immer im Anschluß an das AAGGAG(3')-Ende eines vorher existierenden Clusters integrieren. Dies deckt sich mit den vorher gemachten Beobachtungen, daß (i) U3-Cluster niemals die erste Position in den Clusteraggregaten einnehmen, und (ii) scheinbar freistehenden U3-Clustern das Ende eines verstümmelten Klasse U-Clusters, zumindest aber die Sequenz AGGAG, vorausgeht (vgl. Abschnitt 3.2.3.1.3, Abb.25). Da diese AGGAG-Sequenz an den Übergängen von zwei Clustern nur einmal auftritt, kann sie nicht von den integrierenden Clustern stammen.

Im Gegensatz zu den U3-Clustern ist das U2a-Cluster in den (5')TAG-Anfang eines U2c-Clusters inseriert, wobei das TAG-Trinukleotid dann dem U2a-Cluster vorausgeht. Auch dieser Befund entspricht den oben gemachten Beobachtungen, daß (i) U2a-Cluster immer die erste Position in den Aggregaten einnehmen, (ii) optionale U2a-Einzelcluster in einen TAG-Konsens inserieren und (iii) allen U2a-Clustern die Sequenz TAG vorausgeht.

Die Analyse der optionalen GC-Cluster innerhalb von Clusteraggregaten ergab damit, daß bei der Addition einzelner Cluster sowohl das 3'Ende als auch das 5'Ende eines GC-Clusters als Integrationsort genutzt wird. Die einzelnen Clusterfamilien zeigen bei der Auswahl ihrer Zielsequenzen unterschiedliche Präferenzen; diese werden im folgenden Abschnitt diskutiert.

#3', -2'	TATTTATTAG (TGAACACCTT-4-AGGGAGGAG) (TCTCGTCCA-11-CCATAAGGAG) ----- TAATAAGTATATAGATAAT
#3-1	TATTTATTAG (TGAACACCTT-4-AGGGAGGAG) (TCTCGTCCA-11-CCATAAGGAG) (TGAGGGTCCC-3-AAAGGAG) TAATAAGTATATAGATAAT
#122'	ATTAATAATAATTTATTATTAAAGAGGAG (TGAGGGATCCC-3-GAACCCCGAAGGAG) ----- AAAATAAATTAATAAAAGTT
#122-123	ATTAATAATAATTTATTATTAAAGAGGAG (TGAGGGATCCC-3-GAACCCCGAAGGAG) (TCGGGGACCC-3-GAAGGAG) AAAATAAATTAATAAAAGTT
#62'	TTATATTAATAATTAATAAGAGGAG (TGAGGGACCT-3-GAACCCCGAAGGAG) ----- AAATATTAATAATAATAAT
#62-61	TTATATTAATAATTAATAAGAGGAG (TGAGGGACCT-3-GAACCCCGAAGGAG) (TGAGGGACCC-3-GAAGGAG) AAATATTAATAATAATAAT
#45'	AGAAATAATAATTAATAATAATAATAAG (TCCCGCCCC-2b-CCCGAAGGAG) ----- TATAACGATATAATAATAAT
#45-44	AGAAATAATAATTAATAATAATAATAAG ₁ (TCCCGCCCC-2b-CCCGAAGGAG) (TGAGGGACCC-3-GAAGGAG) TATAACGATATAATAATAAT
#105', 107'	AAATATTAATAATAATAAG (TCCGGCCCC-2a-CCCGAAGGAG) ----- (TGAGGGACC-3-CCCGAAGGAG) TTTAATTTATA
#105-107	AAATATTAATAATAATAAG (TCCGGCCCC-2a-CCCGAAGGAG) (TGAGGGACCC-3-TAAGGAG) (TGAGGGACC-3-CCCGAAGGAG) TTTAATTTATA
#59'	ATTATTTATTTATTTAATAG (TCCGGCCT-12a/3-CGAAGGAG) ----- TTATAAATTTTATTTAACAATAAATTTTAAAT
#59-58	ATTATTTATTTATTTAATAG (TCCGGCCT-12a/3-CGAAGGAG) (TGAGGGACCC-3-GAAGGAG) TTAATAATTATTAATAATAATAATTTTAAAT
#56'	AAATATAAATAAATAAATAAG ----- (TCCGG-2c-GAAGGAG) TTTAATTTCTATTAATAATATATA
#57-56	AAATATAAATAAATAAATAAG (TCCGGCCCC-2a-CCCGAAGGAG) (TCCGG-2c-GAAGGAG) TTTAATTTCTATTAATAATATATA

1 = TCCCGCCCCGGG (-181 bp AAT-reich) -CTCCTTCGGGG

Abb.28 Targetsequenzen optionaler U-Cluster in GC-Cluster-Aggregaten

Zur Legende siehe Abb.24 und Abb.25. Vgl. auch Anhang A.

3.2.3.1.6 FAMILIENBEZOGENE ZUSAMMENFASSUNG ALLER GENANNTEN U-CLUSTER-CHARAKTERISTIKA

Auf Grund ihrer Primärsequenz wurden alle U-Cluster distinkten Familien zugeordnet; sie wurden in Abschn. 3.2.3.1.1 vorgestellt (Abb.20 bis 24). Viele diese Familien zeigten unterschiedliches Verhalten bezüglich ihrer Mobilität und ihrer bevorzugten Zielsequenz (Abb. 24) und damit ihrer Neigung, komplexere GC-Clusteraggregate zu formieren (Abb.25 und 28). Im Folgenden werden die oben gemachten Beobachtungen für die einzelnen Klasse U-Familien gesondert zusammengefaßt.

3.2.3.1.6.1 Familie U1a

Mit einer Ausnahme (TAA) geht allen einzelstehenden U1a-Clustern das Trinukleotid TAG (25 Fälle) oder das erweiterte Targetmotiv TAAAG bzw. TAAAG (2 Fälle) voraus. Bei den drei Tandemaggregaten, an denen U1a-Cluster (#42, #55, #65) beteiligt sind, folgen sie unmittelbar dem (3')AGGAG-Ende von U4-Clustern. Sie besitzen damit im Gegensatz zu den freistehenden U1a-Clustern kein vorausgehendes TAG. Im Übrigen entsprechen diese drei Cluster ihrem Familien-Konsens. Nur in einem der drei Cluster (#42) weicht eine Base, welche an einer weniger konservierten Position steht, von diesem Konsens ab.

Die Analyse der drei optionalen U1a-Cluster (#79, #71, #53) zeigte, daß das einleitende TAG-Trinukleotid (2 Fälle) oder ein verwandtes Oligonukleotid (TAT) in den GC-Sequenzen anderer Hefestämme bereits enthalten ist.

Zusammenfassend kann gefolgert werden, daß das TAG-Trinukleotid kein integraler Bestandteil der Cluster ist, sondern die Zielsequenz ihrer Insertion darstellt (Targetsite-Hypothese). Die drei U4-U1a-Clusteraggregate (#64/#65, #54/#55, #42/#43) könnten durch nachträgliche Insertion der U4-Cluster in den (5')TAG-Anfang der U1a-Cluster erklärt werden. Es kann jedoch nicht ausgeschlossen werden, daß auch das GAGGAG(3')-Ende der U4-Cluster die Integration von U1a-Clustern erlaubt; ihre Integration in das 3'Ende anderer GC-Cluster wurde nicht beobachtet.

3.2.3.1.6.2 Familie U1b

Bei den selteneren U1b-Clustern wurde eine vergleichbare Situation angetroffen. Zwei der vier Cluster (#28,#29) geht ebenfalls das Trinukleotid TAG voraus, bei den beiden übrigen weicht jeweils ein Nukleotid vom TAG-Konsens ab. Im Falle des einzigen optionalen U1b-Clusters (#14) war das TAG-Trinukleotid an der Insertionsstelle im GC-Stamm bereits vorhanden. Die U1b-Cluster bestätigen damit ebenso die Targetsite-Hypothese, ihre Beteiligung an Clusteraggregaten konnte nicht festgestellt werden.

3.2.3.1.6.3 Familie U2a

Mit einer Ausnahme (GAG) geht auch allen 24 U2a-Clustern das TAG-Trinukleotid voraus. Dieses findet sich auch an der Insertionsstelle eines GC-Stamms wieder (#78-). In drei weiteren GC-Stämmen welche zum Vergleich der beiden verbleibenden optionalen U2a-Cluster herangezogen wurden (#63-, #77-, #60-), ist dieses Motiv zu TAAG erweitert oder leicht variiert (TAA bzw. TGAG). Dies könnte Mutationseignisse an der GC-Sequenz widerspiegeln und damit ein Hinweis dafür sein, daß die Insertion dieser optionalen U2a-Cluster länger zurückliegt als jene der beiden U1-Familien. Für ein höheres phylogenetisches Alter dieser Familie spricht auch, daß die Sequenz der U2a-Cluster weit mehr von ihrem Konsens abweicht als diejenige der beiden erstgenannten.

Neun der U2a-Cluster (40%) sind an der Ausbildung von Clusteraggregaten beteiligt, wobei sie immer an erster Stelle innerhalb des Aggregates stehen und von U3-Clustern (6 Fälle) oder U2c-Clustern (3 Fälle) gefolgt werden. Die Analyse eines optionalen U2a-Clusters (#57) ließ die Insertion des Clusters in den (5')TAG-Anfang eines U2c-Cluster erkennen. Die enorme Konservierung des TAG am 5'Anfang der U2a-Cluster, ihr häufiges Auftreten am 5'Anfang komplexer Clusteraggregate und die Insertion optionaler U2a-Cluster in die Konsenssequenz TAG belegen damit, daß das Trinukleotid TAG auch für U2a-Cluster die bevorzugte Insertionssequenz darstellt, wobei diese Cluster eine auffällige Neigung zur Insertion in den 5'Anfang bereits existierender Cluster zu haben scheinen.

3.2.3.1.6.4 Familie U2b

Nur einem der sechs U2b-Cluster (#83) geht ebenfalls das TAG-Trinukleotid voran, die fünf restlichen folgen den erweiterten Motiven TAAG, TAAAG, TAGAG. Entsprechend finden sich diese erweiterten Motive auch in den GC-Sequenzen (#45', #70-) der beiden optionalen Cluster dieser Familie wieder, wobei in einem der Fälle das an einer Position abweichende TAAGG-Oligonukleotid gefunden wurde. Drei der Cluster wurden in den homologen Startsequenzen der mtDNA-Replikation gefunden (ori8, #70; ori4, #83; ori6, #113). Der kein GC-Cluster enthaltende Replikationsstart (ori5) besitzt an der entsprechenden Stelle wiederum das erweiterte Targetmotiv TAAAG. Auch diese Cluster bestätigen damit die Targetsite-Hypothese, wobei die Vorliebe der Cluster für erweiterte Targetmotive auffällt.

3.2.3.1.6.5 Familie U2c

Nur vier der 10 U2c-Cluster treten als einzelstehende Cluster auf, diesen gehen wiederum die typischen Targetmotive TAG (3 Fälle) und TAAG (1 Fall) voran. Eines der sieben U2c-Cluster der Clusteraggregate steht am Anfang des Aggregats. Die sechs restlichen folgen unmittelbar an das AAGGAG Ende der vorangehenden U3- oder U2a-Cluster. Entsprechend besitzen diese U2c-Cluster kein TAG-Trinukleotid. In einem der letzteren Fälle ist das U2a-Cluster optional (#57), in diesem Fall folgt das U2c-Cluster unmittelbar dem üblichen TAG-Motiv. Auch diese Cluster bestätigen damit die Targetsite-Hypothese.

Es konnte kein optionales U2c-Cluster entdeckt werden. Da diese extrem kurzen Cluster nur die Anfangs- und Endsequenzen der U2a-Cluster besitzen, besteht die Möglichkeit, daß sie in ihrer heutigen Form die Fähigkeit zur Mobilität verloren haben. Sie würden damit spezifische Deletionsprodukte von U2a-Clustern (bzw. in einzelnen Fällen auch von U1a-Clustern) darstellen. Wären auch eines der Clusterenden von Deletion betroffen worden, so (i) wären sie im einzelstehenden Fall nicht als GC-Cluster erkannt worden, (ii) hätten sie nicht die nötigen Targetsequenzen zum Aufbau der Clusteraggregate zur Verfügung gestellt.

3.2.3.1.6.6 Familie U3

Diese Familie unterscheidet sich von den restlichen U-Clustern insofern, als nur einem der 34 U3-Cluster das TAG-Trinukleotid vorausgeht (#81). Im übrigen schließen U3-Cluster unmittelbar an die Sequenz AGGAG an, welche das typische Endmotiv der Klasse U darstellt (30 Fälle). Entsprechend ist der überwiegende Teil aller U3-Cluster (23 Fälle) Bestandteil von Clusteraggregaten, wobei sie regelmäßig dem AAGGAG Ende anderer Cluster folgen (2 Ausnahmen).

Da dieses AGGAG auch 10 der 11 U3-Clustern vorausgeht, welche außerhalb von Aggregaten angetroffen werden, muß diese Sequenz als Zielsequenz der U3-Cluster betrachtet werden. Wie die fünf Beispiele optionaler U3-Cluster (immer in Aggregaten) zeigen, ist diese Sequenz an den Insertionsstellen der GC-Stämme bereits vorhanden. Die Cluster folgen damit im Prinzip ebenfalls der Targetsite-Hypothese, wobei als 'Targetsite' der U3-Cluster die Sequenz AAGGAG und damit das 3'Ende eines Klasse U-Clusters angesehen werden muß. Wie die einzelstehenden U3-Cluster zeigen, muß in dem vorausgehenden Clusterfragment nur noch die Endsequenz AGGAG erhalten sein. Ob die Degeneration dieser vorangehenden Cluster vor oder nach der Insertion der U3-Cluster stattfand, kann nicht entschieden werden. Die Insertion optionaler U3-Cluster konnte nur in die 3'Enden gut erhaltener GC-Cluster beobachtet werden.

Die U3-Targetsequenz (AAGGAG) ist auch am Ende der U3-Cluster (wie bei allen Klasse U-Clustern) zu finden. Daher könnte bei der Insertion von U3-Clustern sowohl das AG-Dinukleotid, wie dies bei den übrigen U-Clustern wahrscheinlich ist, als auch ein längeres Oligonukleotid der Targetsequenz (z.B. AAGGAG) verdoppelt werden. Da im letzteren Falle sowohl das dann neu zu definierende Clusterende als auch die verdoppelten Targetnukleotide von den restlichen U-Clustern abweichen würde, müßte die Zuordnung der Cluster zur Klasse U neu überdacht werden.

Interessant ist in diesem Zusammenhang, daß das einzige U3-Cluster (#81), welches anstelle dem AAGGAG-Motiv dem TAG-Trinukleotid folgt, zugleich das einzige U3-Cluster ist, welches nicht durch das AAGGAG-Motiv beendet wird. Hier findet sich auch am Ende Clusters das TAG-Trinukleotid wieder.

3.2.3.1.6.7 Familie U4

Alle vier U4-Cluster beginnen mit dem TAG-Trinukleotid. In Analogie zu U1- und U2-Clustern kann dieses Trinukleotid daher als Targetsequenz dieser Cluster vermutet werden. Da die Cluster immer U1a-Clustern (3 Fälle) oder einem damit verwandten imperfekten Cluster (1 Fall) vorangehen, kann vermutet werden, daß sie vorzugsweise in das (5')TAG Ende von U1a-Clustern integrieren und damit eine komplexere Targetsequenz aufweisen. Bei den beiden optionalen Clustern dieser Familie fehlt in den GC--Sequenzen jeweils das gesamte U4-U1a-Tandem, so daß sie wenig Aufschluß über den Insertionsmodus der U4-Cluster liefern. In einem der beiden Fälle wurde die Nukleotidsequenz am Rande des GC-Cluster-Tandems nicht publiziert, so daß der Insertionsort nicht eindeutig bestimmen werden konnte. Bei dem zweiten U4-U1a-Tandem konnte das TAG-Trinukleotid an der Insertionsstelle der GC--Sequenz (#64-) gefunden werden. Wie alle Klasse U-Cluster werden U4-Cluster durch das AG-Dinukleotid beendet.

3.2.3.2 DIE KLASSE G

3.2.3.2.1 NUKLEOTIDSEQUENZEN DER G-CLUSTER

Die GC-Cluster der Klasse G bilden nur eine kleine, aus fünf Clustern bestehende, homogene Familie, welche sich in ihrer Primärsequenz insbesondere auch an den beiden Clusterenden von den restlichen GC-Clustern unterscheiden. Alle diese Cluster sind in den sog. Replikationsstart-Sequenzen inseriert (ori4, ori6, ori7, (128); ori8, ori1*,(34)). Da die Replikationsstart-Sequenzen (ori1, ori2, ori3, ori5) mit den vorher genannten homolog sind (128), jedoch keine Klasse G-Cluster enthalten, können alle Cluster dieser Familie als optional angesehen werden.

Abb.29 gibt den Bereich der Replikationsstart-Sequenzen (ori1 bis 8), der sich durch die Anwesenheit eines G-Clusters unterscheidet, wieder.

G		AAATATATAAGT	CCCGGTTTCTTACGAAACCGGGACCTCGGAGACGT	AATAGGGGG
#69	ORI8	AAATATATAAGT	CCCGGTTTCTTACGAAACCGGGACCaCGGAGACGT	AATAGGGGG
#87	ORI7	AAATATAT-AG-	--CGGTTTCTTACGAAACCGGGACCTCGGAGACGT	AATAGGGGG
#112	ORI6	AAATATAT-AGT	CCCGGTTTCTTACGAAACCGGGACCTCGGAGAAGT	AATAGGGGG
#133	ORI4	AAATATAT-AGT	CCCGGTTTCTTACGAAACCGGGACCTCGGAGACGT	AATAGGGGG
#47	ORI1	AAATATATAAGT	CCCGGTTTCTTACGAAACCGGGACCTCGGAGACGT	AATAGGGGG
#47 ¹	ORI1	AAATATATAAGT	-----	AATA ¹ GGGG
	ORI2	AAATATATAAGT	-----	AATA ¹ GGGG
	ORI3	ggATATATagGT	-----	ATTAGGGGG
	ORI5	AAATATATAAGT	-----	AATAGGGGG

Abb.29 G-Cluster

¹ = AATTAAGTTTATA (siehe Text).

Ohne das einleitende Motiv (AGT) ist die Konsenssequenz der G-Cluster 35 bp lang, sie enthält zu 60% G+C-Nukleotide. Eines der Cluster (#47) wird in der entsprechenden Sequenz (#47-) eines anderen Stammes nicht angetroffen.

3.2.3.2.2 TARGETSEQUENZEN DER G-CLUSTER-INTEGRATION

Sowohl der Vergleich des optionalen Clusters (#47) mit seiner GC-Sequenz (#47-) (vgl. Abb.29), als auch der Vergleich clusterfreier und Cluster enthaltender Replikationsstart-Sequenzen zeigt, daß G-Cluster an die Sequenz AGT anschließen. Das Trinukleotid AGT ist damit Bestandteil der Targetsequenz von G-Clustern.

Das den G-Clustern vorausgehende Dinukleotid GT findet sich auch am Ende der Cluster wieder. Somit könnte die Integration von G-Clustern (ähnlich wie bei U-Clustern) eine Verdoppelung zweier Nukleotide der Insertionsstelle bewirken.

Da alle G-Cluster in sog. Replikationsstart-Sequenzen lokalisiert sind, können bis zu 300 bp, welche in diesen Sequenzen homolog sind, als Targetsequenzen der G-Cluster in Frage kommen. Für die Annahme einer komplexen Targetsequenz spricht auch, daß die Cluster keine Aggregate ausbilden. Im Gegensatz zur der kurzen Targetsequenz (TAG) der Klasse U, würde eine längere Targetsequenz durch die Insertion eines GC-Clusters zerstört werden; sie könnte daher nicht zur Integration weiterer Cluster genutzt werden. Im mt Genom von *Sacch.cerevisiae* befindet sich die Klasse G-Targetsequenz in beiden möglichen Orientierungen. Die Orientierung der G-Cluster wird in allen Fällen durch ihre Targetsequenz determiniert.

3.2.3.3 DIE KLASSE V

3.2.3.3.1 NUKLEOTIDSEQUENZEN DER V-CLUSTER

Diese Klasse enthält nur vier Mitglieder, welche eine homogene Familie bilden: ihre Sequenzen sind in Abb.30 wiedergegeben.

V		CCCCGCGGGCGCCAATCCGGTTGTTACCGGATTGGTCCCGCGGGG
#37	*	ccCCCCGCGGGCGCCAaCCCCGTTGcTCACCGGGTTGGTCCCaCGGGGg
#121		CCCCGCGGGCGCCAATCCGNNNNNNNCCGGATTGGTCCCGCGGGG
#128		CCCCGCGGGCGCCAATCCGGTTGTTACCGGATTGGTCCCGCGGGG
#129	*	CCCCGCGGGCGCCAATCCGGTTGTTACCGGATTGGTCCCGCGGGG

Abb.30 V-Cluster

Zur Legende vgl. Abb.20 (siehe Text).

Zwei der Cluster sind identisch, ein drittes (#121) unterscheidet sich scheinbar durch eine Lücke von 7 bp, sie ist durch den Verlust eines HpaII-Fragmentes bei der Sequenzierung der Cluster begründet (118, 67). Die Konsenssequenz der G-Cluster ist 46 bp lang, ihr G+C-Gehalt beträgt 73,9%. Sie zeigt keine signifikante Homologie zu derjenigen anderer GC-Cluster. Die neun terminalen Nukleotide bilden ein Palindrom. Zwei der Cluster (#37, #129) sind optional, wobei mir die GC-Sequenz des etwas varianten Clusters #37 nicht vorliegt.

3.2.3.3.2 TARGETSEQUENZEN DER V-CLUSTER-INTEGRATION

Die Targetsequenz der V-Cluster weicht erheblich von jener der übrigen GC-Cluster-Klassen ab. Bei drei der vier V-Cluster sind an beiden Seiten der Cluster je 11 flankierenden Nukleotide identisch (vgl. Abb.31). An der Targetsequenz (#129-) eines optionalen V-Clusters (#129) sind 20 dieser Nukleotide bereits vorhanden, die Insertion des Clusters erfolgt in ein AA-Dinukleotid in der Mitte der Targetsequenz. Randsequenzen und Clustersequenzen des vierten V-Clusters (#37) sind etwas weniger konserviert.

	<u>Flanke</u>	<u>V-Cluster</u>	<u>Flanke</u>
#37	-> att AATtA-ATAAA	(CCC => GGG)	ttaAaTAATat att ->
#121	<- tta AATAATATAAA	(CCC => GGG)	AATATTAATAA att <-
#128	-> aat AATAATATAAA	(CCC => GGG)	AATATTAATAA taa ->
#129	-> tta AATAATATAAA	(CCC <= GGG)	AATATTAATAA aat ->
#129 ⁻	-> tta AATAATATAAA	(-----)	--TATTAATAA aat ->

Abb.31 Targetsequenzen der V-Cluster

'->' gibt die Orientierungen der Sequenzen (Flanken und Cluster) im Genom wieder (-> = Transkriptionsrichtung). '=>' zeigt die Polarität der GC-Cluster; die vollständigen Sequenzen der Cluster wurden in Abb.30 wiedergegeben (=> Richtung). Die untere Zeile zeigt die Targetsequenz (#129⁻) des optionalen Clusters (#129).

Wie in Abb.31 dargestellt zeichnen sich V-Cluster damit durch folgende Eigenschaften aus:

- (i) Die Orientierung der Cluster wird nicht durch die Orientierung ihrer Targetsequenz bestimmt. Beliebige Kombinationen werden gefunden (->=>->, -><=>, <=><->).
- (ii) Im GC--Stamm ist die Targetsequenz (#129⁻) 20 bp lang, das Cluster inseriert in der Mitte dieser Zielsequenz in ein AA-Dinukleotid.
- (iii) Nach der Insertion des Clusters ist dieses Dinukleotid an beiden Enden des GC-Clusters vorhanden (#129). Es ist unwahrscheinlich, daß das AA-Dinukleotid integraler Bestandteil des inserierenden V-Clusters ist, da es in diesem Fall bei entgegengesetzter Orientierung des Clusters in ein TT-Dinukleotid umgewandelt würde. Dies wurde in der gut konservierten Sequenz (#128) nicht beobachtet (in Abb.31 fett gedruckt).

Auf die Verdoppelung des AA-Dinukleotids wurde in verschiedenen Arbeiten der Arbeitsgruppe R.Butow (19, 56) bereits hingewiesen.

3.2.3.4 DIE KLASSE I

In dieser imperfekten Klasse sind die restlichen nicht näher klassifizierten GC-Cluster zusammengefaßt (vgl. Abb.32). Sie setzt sich aus folgenden GC-Cluster-elementen zusammen:

(i) Fusionsprodukte unterschiedlicher Cluster, welche entweder durch Rekombination einzelner Cluster oder durch unspezifische Deletion innerhalb eines Clusteraggregates entstanden sind. Hierzu zählt das optionale 'Mini-Insert' der 21S rRNA (#137) dessen 5'Hälfte durch den Anfang eines U1-Clusters gebildet wird, die 3'Hälfte des Clusters stellt das Ende eines U4-Clusters dar. Das Cluster (#59) beginnt mit einem U2a-Cluster und geht etwa in der Sequenzmitte abrupt in ein U3-Cluster über. Eine noch weiter degenerierte Form des Clusters (#59') wurde in einem anderen Hefestamm gefunden.

(ii) Basenaustausche und Deletionen innerhalb eines Clusters, wobei in einigen Fällen typische Sequenzbereiche erhalten blieben, so daß die Herkunft des Clusterfragments noch vermutet werden kann. So stellen die Cluster (#99, #111, #12, #81') verstümmelte Formen eines U3-Clusters dar. Letzeres (#81') ist in einem anderen Hefestamm als vollständiges U3-Cluster (#81) erhalten geblieben. Cluster #100 ist zu U1-Clustern homolog. Die Cluster #23 und #2, welche beide Bestandteile von Clusteraggregaten darstellen, ähneln sowohl U1-, als auch U2-Clustern; ihre Herkunft kann nicht genauer angegeben werden. Clusterfragment #19 erinnert an U4-Cluster und Cluster #74 ist ein reduziertes V-Cluster. Die Herkunft der verbleibenden Cluster Fragmente konnte nicht entschieden werden.

(iii) GC-Cluster Fragmente mit weniger als 12 G+C-Nukleotiden (#46) wurden hier nur aufgenommen, wenn sie Bestandteile komplexerer Clusteraggregate darstellten. Betrachtet man G+C-Tetranukleotide bereits als GC-Cluster, so würde sich diese Liste um etwa 35 Klasse I-Cluster erweitern.

```
#2 (#2') TCTCGGTCCAACCGGGACCCATAAGGAG
#12 TAACCGGGCGGGGACCCCGCCACGGAAGGCCCCGAAG
#19 ATGTGAATCCTGCTAGGAGGAG
#23 CCCTTCGGAACCCCGAATCTCGTTGAGGGT
#32 TAGTCCCACTCGCGTAACGGTGTCCGC
#46 CTCCTTCGGGG
#59 TAGTCCGGCCCCCGCCGCGGG/AACCGTACCCCTTTTTAAGAAGGAG
#59' TAGTCCGGCC/GAACCCGAAGGAG
#74 CCCACGAGGGCCACACATGTGTGGCCTCGCGGGG
#75 CCCACCCCCATGCGAAGCATGGGGGGGGTATAAGTATGGACAATCCGCAGGA
#81' TAGGACTCCCGAAGTAACACTACGGGAGGACCGACCCTG
#88 CTGGGGACCACCCCCCACC
#92 CACTGCGGGGTTGGCGCCC
#93 TAGTGTGGGGGAGGGGAGGAG
#99 TAGTCCGTGGGATCGAACCCCTTTTTTATTTAATATTTAATATTTAAAGAAGGAA
#100 TAAGTTTTGGATACCAGGCCAAAGACCGGAATCCCAAAAGGAG
#111 TAGGGTGAGTAGGACGCCGATGAGAGAGGGATCCCCC
#131 TAGTTCGGTTCCGATATGGACCAAACTCCTAATGGAG
#137 TAAGTTCGGAACCCCGAAGAGGG/
      GTTCACACCTATTAAAAATAGGTGTGAACCCCCACCGAGAGAGGAG
```

Abb.32 I-Cluster

Sequenzbereiche, welche Teilen vollständiger GC-Clustern eindeutig zugeordnet werden können, sind durch '/' begrenzt. Vgl. auch Anhang A (siehe Text).

3.2.4 DIE MERKMALE DER DREI HAUPTKLASSEN IM VERGLEICH

Alle GC-Cluster sind G+C reiche, repetitive Sequenzen, welche vielfach in identischer oder ähnlicher Form im mt Genom von *Sacch.cerevisiae* angetroffen werden. Die Sequenzen der Vertreter aller perfekten Klassen (U, G, V) stellen unvollständige Palindrome dar: sie können stabile 'Hairpinstrukturen' ausbilden.

Die Cluster werden an beiden Enden von klassen-spezifischen Dinukleotiden (AG, GT, AA) begrenzt. Die selben Dinukleotide werden auch an den Insertionsstellen der optionalen Cluster (GC-Sequenzen) gefunden. Es ist daher wahrscheinlich, daß diese Dinukleotide durch die Verdoppelung des Targetdinukleotids entstanden (vgl. Abb.33).

Klasse U	NNTAG (TNN==>AGG) AGNN
	NNTAG ----- --NN
Klasse G	...AAATATATAAGT (CCC==>ACG) GTAATAGGGGG...
	...AAATATATAAGT ----- --AATAGGGGG...
Klasse V	NNAATAATATAAA (CCC<=>GGG) AATATTAATAANN
	NNAATAATATAAA ----- --TATTAATAANN

Abb.33 Dinukleotid-Verdoppelung der drei GC-Cluster-Klassen

GC-Cluster sind nur durch ihre Anfangs- und Endsequenzen dargestellt. Pfeile geben ihre möglichen Orientierungen zur Targetsequenz wieder. Die vollständige Targetsequenz der Klasse G könnte auch durch einem größeren Teil der Replikationsstart-Sequenz gestellt werden (siehe Text).

Im Falle der V-Cluster, deren Orientierung von ihrer Targetsequenz unabhängig ist, kann gezeigt werden, daß diese terminalen Dinukleotide nicht an der Inversion der Cluster beteiligt sind, sondern die Orientierung der Targetsequenz beibehalten und damit nicht Bestandteile der Cluster darstellen können (vgl. Abschn. 3.2.3.3.2).

Die bezüglich zur Targetsequenz beliebige Orientierung der V-Cluster könnte durch die Symmetrie der Clusterenden bedingt sein. Je 9 bp an den beiden Enden der V-Cluster stellen zueinander palindrome Sequenzen dar. Die Endsequenz ist daher von der Orientierung der Cluster unabhängig. Die Enden von U- und G-Clustern sind dagegen unsymmetrisch. Möglicherweise wegen ihrer unsymmetrischen Endsequenzen wird die Orientierung dieser Cluster stets durch die Orientierung ihrer Targetsequenzen determiniert.

Die kurze Targetsequenz (3 bis 6 bp) der U-Cluster bleibt nach Insertion der Cluster erhalten. Entsprechend neigen viele dieser Cluster zur Ausbildung komplexer Tandemaggregate. Die weite Verbreitung dieser Cluster könnte ebenfalls durch die einfache (und damit häufig auftretende) Struktur ihrer Targetsequenz bedingt sein. Die wesentlich komplexeren Targetsequenzen der seltenen Cluster-Klassen G und V werden durch die Insertion der Cluster zerstört, ihre Aggregation wurde daher nicht beobachtet.

Abb.34 stellt die genannten Charakteristika der drei GC-Cluster-Klassen tabellarisch gegenüber.

Klasse	Targetsequenz	Verdoppl.	Orientierung	Anzahl	Tandems
U	T(A)AG (AAGGAG)	AG	Target	113	+
G	...T(A)AGTAATA...	GT	Target	5	-
V	AATAATATAAATATTAATAA	AA	beliebig	4	-

Abb.34 Charakteristika der drei GC-Cluster-Klassen im Vergleich

Die Insertionsorte innerhalb der Targetsequenzen sind hervorgehoben. Weiterhin sind gegenübergestellt: die bei der Insertion der Cluster vermutlich verdoppelten Dinukleotide, die Orientierung der Cluster bezüglich ihrer Targetsequenz, die Anzahl der bisher vollständig sequenzierten Cluster und die Fähigkeit der Cluster Tandemaggregate auszubilden.

3.2.4.1 EXKURS: VERGLEICH DER GC-CLUSTER-SYSTEMATIK MIT DEN ERGEBNISSEN VON ZAMAROCZY & BERNARDI

Im selben Zeitraum, in dem die vorliegende Arbeit entstand, wurden von Zamaroczy & Bernardi systematische Untersuchungen an GC-Clustern vorgenommen. Diese wurden während der Endfassung meiner Arbeit publiziert (129, 130). Beide Untersuchungen wurden unabhängig von einander durchgeführt. Meine Arbeit bestätigt diejenige von Zamaroczy & Bernardi im Wesentlichen. Auf Grund eines unterschiedlichen methodischen Ansatzes und insbesondere durch meine Analyse optionaler GC-Cluster und deren Targetsequenzen, sowie der Organisation komplexer GC-Clusteraggregate, gehen meine Ergebnisse in wichtigen Punkten über die von Zamaroczy & Bernardi gefundenen Daten hinaus. Es erscheint daher an dieser Stelle geboten, beide GC-Cluster-Systeme mit einander zu vergleichen.

Die Nomenklatur von Zamaroczy & Bernardi orientiert sich an kurzen **Sequenzblöcken**. Diese wurden mit 'a', 'b', 'c', 'i', 'm', 'q' und 't' bezeichnet (129). Viele dieser Sequenzblöcke werden in ähnlicher Form in den unterschiedlichen GC-Cluster-Typen angetroffen. Ausgehend von der Verteilung dieser Blöcke innerhalb der Cluster, teilten die Autoren die GC-Cluster in sieben Familien ein (a1, a2, a3, a4, c, n, v). Die wechselseitige Verwandtschaft der einzelnen GC-Cluster-Familien ergab sich dabei aus dem gemeinsamen Besitz gleicher oder ähnlicher Sequenzblöcke.

Meine Analyse orientierte sich im Gegensatz dazu an der kleinstmöglichen Anzahl von Konsenssequenzen, welche nötig sind, die **vollständigen GC-Clustersequenzen** zu beschreiben. Unvollständig sequenzierte Cluster schloß ich daher von der Analyse aus. Die Entwicklung geeigneter Software (vgl. Abschn. 2.2.2) erlaubte mir, die Konsenssequenzen der Cluster zu ermitteln und die Signifikanz einzelner Abweichungen von diesen Konsenssequenzen zu beurteilen. Außer an den Primärsequenzen der Cluster und der Sequenzverwandtschaft ihrer Konsenssequenzen orientiert sich die von mir entwickelte Systematik zusätzlich an folgenden, von mir erstmals systematisch analysierten, GC-Cluster-Kennzeichen:

- (i) Targetsequenzen der GC-Cluster-Integration.
- (ii) Aggregationsverhalten der Cluster

(iii) Wahrscheinliche Verdoppelung von Targetnukleotiden bei der Cluster-Integration.

(iiii) Orientierung der Cluster bezüglich ihrer Targetsequenzen.

Auf Grund dieser Merkmale konnte ich die Cluster in 10 Gruppen untergliedern. Da sich manche der Gruppen in nur wenigen, andere in vielen der genannten Kennzeichen unterscheiden, stellt meine Klassifikation ein hierarchisches System mit Klassen, Familien und Unterfamilien dar.

Trotz der unterschiedlichen Einteilungskriterien wurde der Großteil aller GC-Cluster von Zamaroczy & Bernardi ebenso gruppiert, wie ich dies in meiner Arbeit tat. Abb.35 stellt beide Gruppierungssysteme gegenüber.

Die von Zamaroczy & Bernardi mit a1, a2 bzw. a3 bezeichneten Familien entsprechen gemäß meiner Nomenklatur Klasse U-Clustern. Die beiden Familien a1 bzw. a2 sind bei mir in die Unterfamilien U1a, U1b bzw. U2a, U2b untergliedert (vgl. Abb.20 bzw. Abb.21).

Die a3-Familie entspricht den U3-Clustern in meiner Nomenklatur.

Die mit 'a4' bezeichnete Familie stellt keine homogene GC-Cluster-Gruppe dar. Sie enthält alle GC-Cluster, welche ein oder mehrere kurze Sequenzmotive mit Clustern der a-Familien gemeinsam haben (129). Ein Teil dieser Cluster wurde von mir der imperfekten Klasse I zugeordnet, wodurch zum Ausdruck kommt, daß es sich hierbei um durch Mutation verstümmelte Cluster der übrigen Familien handeln könnte (vgl. Abschn.3.2.3.4).

Einige mit 'a4' bezeichnete kurze GC-Cluster, welche häufig in Clusteraggregaten angetroffen werden, sind wegen ihrer hohen Sequenzverwandschaft mit U2-Clustern in meiner Nomenklatur in der Unterfamilie U2c zusammengefaßt (vgl. Abb.21). Andere a4-Cluster unterscheiden sich gemäß meinen Untersuchungen sowohl in ihren Clusterenden als auch in ihren Targetsequenzen und in den bei ihrer Integration vermutlich verdoppelten Dinukleotiden erheblich von den übrigen GC-Clustern. Diese wurden von mir daher einer eigenen Klasse (Klasse G) zugeordnet (vgl. Abb.29).

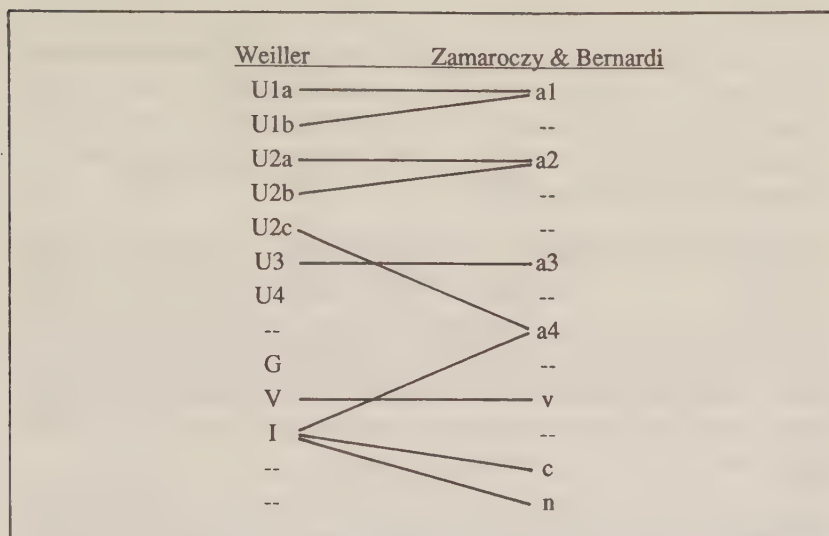


Abb.35 Vergleich der GC-Cluster-Systematik mit einem von Zamaroczy & Bernardi parallel entwickelten Einteilungssystem

Die Abb. vergleicht die in beiden Systemen gruppierten Cluster und deren Bezeichnungen. Die Cluster der heterogenen Familien a4, c und n wurden von Zamaroczy & Bernardi teilweise abweichend gruppiert (siehe Text).

Die Klasse V (Var-Cluster) entspricht der v-Familie von Zamaroczy & Bernardi. Einige kürzere mit V-Clustern verwandte Cluster-(Fragmente) wurden in meiner Arbeit der imperfekten Klasse I zugeordnet. Da V-Cluster sich in allen o.g. Kennzeichen von den übrigen GC-Clustern unterscheiden, bilden sie in meiner Systematik eine eigene Klasse.

Bei den Clustern, welche von Zamaroczy & Bernardi in den Familien 'c' bzw. 'n' bezeichnet wurden, handelt es sich meist um sehr kurze (≥ 4 bp) G+C-reiche Sequenzmotive. Die c-Familie zeichnet sich dabei durch eine Serie von vier und mehr C-Nukleotiden aus. Ein ähnliches Motiv findet sich auch in Cluster-C der Replikationsstartsequenzen (129) bzw. in U3-Clustern wieder. Die kurzen Formen dieser Cluster wurden in meiner Untersuchung aus methodischen Gründen von der Analyse ausgeschlossen, die längeren Cluster (≥ 12 G+C-Nukleotide) sind in meiner Nomenklatur der Klasse I zugeordnet.

Mein Interesse galt im Besonderen der Mobilität der Cluster. Daher sequenzierte ich acht optionale GC-Cluster. Durch den Vergleich dieser und weiterer in der Literatur zu findender optionaler GC-Cluster (GC⁺-Sequenzen) mit den GC-Cluster freien Allelen anderer Hefestämme (GC⁻-Sequenzen) konnte ich die Targetsequenzen der Cluster bestimmen.

So konnte ich erstmals zeigen, daß sowohl die Sequenz TAG, welche regelmäßig am 5'Anfang von U1-, U2- und U4-Clustern auftritt, als auch die Sequenz AAGGAG, welche stets am 5'Anfang der U3-Cluster zu finden ist, am potentiellen Integrationsort in den GC⁻-Stämmen bereits enthalten ist (vgl. Abb.24 und Abb.28). Diese am 5'Ende der Cluster konservierten Sequenzen bilden daher keinen integralen Teil der GC-Cluster sondern die bisher nicht als solche erkannten Targetsequenzen, in die die Cluster integriert sind.

Die Entdeckung der Targetsequenzen erlaubte mir erstmals, die Endpunkte der Clustersequenzen zu bestimmen. Diese Ergebnisse konnte ich sowohl durch die Analyse der Organisation komplexer GC-Clusteraggregate, als auch durch die Sequenzierung polymorpher GC-Clusteraggregate bestätigen (vgl. Abschn. 3.2.3.1.2 bis 3.2.3.1.5).

Meine Definition der Klasse U GC-Clusterenden unterscheidet sich daher von derjenigen von Zamaroczy & Bernardi in sofern, als ich die Sequenz 5'TAG bzw. 5'AAGGAG nicht als Bestandteil der GC-Cluster bezeichne. In der Nomenklatur von Zamaroczy & Bernardi finden sich die Targetsequenzen TAG bzw. AAGGAG in den Clusterblöcken 't' (Sequenz ATAGTT) bzw. 'i' (Sequenz GAAGGAG) wieder. Durch den systematischen Vergleich der G-Cluster mit ihren Targetsequenzen konnte ich auch die Endpunkte dieser Clusterklasse neu bestimmen (vgl. Abb.29).

Durch die Analyse der einzelnen Bestandteile der Clusteraggregate stieß ich auf eine weitere GC-Clusterfamilie (U4) welche bisher nicht beschrieben wurde (vgl. Abb.23). Diese Cluster werden ausnahmslos in Kombination mit anderen GC-Clustern (U4-U1a, U4-I1a-U3) gefunden (vgl. Abb.25).

Zusammengenommen erlauben meine Ergebnisse Rückschlüsse auf den Integrationsmechanismus der Cluster. Die Mobilität der Cluster, mögliche Mechanismen ihrer Integration und damit möglicherweise einhergehende Rearrangements (Deletionen) des mt Genoms werden in den folgenden Abschnitten behandelt.

3.2.5 MOBILITÄT VON GC-CLUSTERN

3.2.5.1 HINWEISE FÜR DAS MOBILE VERHALTEN DER CLUSTER

GC-Cluster zeichnen sich durch bemerkenswerte Mobilität aus. Die wichtigsten Hinweise hierfür sind im Folgenden zusammengefaßt.

(i) Die Cluster werden in identischer oder ähnlicher Form in der mt DNA gefunden (repetitive Sequenzen). Ihre weite Verbreitung legt ihre Ausbreitung in der Evolution nahe.

(ii) Einzelne Cluster sind auch in Genen inseriert (VAR1, 15S RNA, 21S RNA, ORF2, ORF3, a15b). In homologen Genen anderer Organismen werden sie nicht gefunden. In einigen Fällen werden die Leserahmen dabei gestört. Die Evolution dieser GC-Cluster war daher unabhängig von der Evolution der Gene, in denen sie angetroffen werden.

(iii) Die Analyse unterschiedlicher Hefestämme erbrachte eine Vielzahl optionaler Cluster. Hieraus kann gefolgert werden, daß die Insertion (oder Exzision) dieser optionalen Cluster erst nach der Aufspaltung der einzelnen Stämme stattfand.

(iii) An den Insertionsstellen optionaler GC-Cluster wurden für jede Clusterklasse typische Targetsequenzen gefunden; diese legen sequenz-spezifische Insertions-mechanismen nahe.

(iiii) Die palindrome Organisation der Cluster zusammen mit ihren kurzen (2 bp) terminalen direkten Sequenzwiederholungen, welche ihren verdoppelten Targetsequenzen entsprechen, erinnern an prokaryontische IS-Elemente. Beide inserieren bevorzugt in leicht aufschmelzbare DNA-Bereiche. Als solche sind sowohl A+T-reiche Sequenzen, als auch die Ränder von GC-Clustern anzusehen, da die Cluster durch ihre ausgeprägte Fähigkeit, 'Hairpinstrukturen' auszubilden, den DNA-Doppelstrang destabilisieren.

Jedes der GC-Cluster sollte daher potentiell als optionales Element auftreten können.

3.2.5.2 HÄUFIGKEIT OPTIONALER GC-CLUSTER

Codierende Sequenzen sind bisher auf großes Interesse gestoßen: sie wurden in unterschiedlichen Hefestämmen untersucht. Entsprechend wurden hier auch die ersten optionalen Cluster gefunden. (15S RNA, 21S RNA, VAR1). Da verschiedene Hefestämme in ihren Restriktionsmustern bezüglich der in den Clustern schneidenden Enzyme HpaII (CCGG) und HaeIII (GGCC) weit mehr variieren als bei anderen Restriktionsenzymen, war anzunehmen, daß sich viele dieser Stämme insbesondere durch ihre GC-Cluster unterscheiden.

In den verfügbaren mt Sequenzen konnte ich 53 GC-Cluster finden, bei denen die entsprechenden Genombereiche in wenigstens zwei unterschiedlichen Hefestämmen sequenziert waren. 29 dieser Cluster (55%) wurden in wenigstens einem der Stämme nicht gefunden. Meist standen nur zwei unterschiedliche Stämme zum Vergleich. Die Wahrscheinlichkeit, ein gegebenes Cluster auch in einen zweiten Stamm anzutreffen, war etwa 50%. Bei einer genügenden Anzahl vergleichbarer Stämme sollte daher jedes GC-Cluster ein optionales Element darstellen können.

3.2.5.3 INSERTION VERSUS DELETION

Es kann nicht ausgeschlossen werden, daß einige dieser GC+/GC⁻-Sequenzpolymorphismen durch Deletion eines GC-Clusters entstanden sind. Da unspezifische Deletion jedoch nicht exakt die Ränder der Cluster betreffen würde, wie dies in fast allen Fällen zu beobachten ist, müßte dabei ein sequenz-spezifischer Exzisionsmechanismus gefordert werden.

Das häufige Auftreten identischer Cluster an unterschiedlichen Orten des Genoms belegt ihre Ausbreitung während der Evolution. Ihre Ausbreitung und ihr optionales Vorkommen auch in codierenden Bereichen (Genen), deren Evolution derjenigen der GC-Cluster vorausging, lassen sich nur durch Insertion von GC-Clustern erklären. Sollte ein Mechanismus existieren, durch welchen GC-Cluster ihr Target sequenz-spezifisch verlassen können, so könnte dieser durch die Umkehrung des Insertionsmechanismus am einfachsten erklärt werden.

3.2.6 TRANSPOSITIONSMECHANISMEN

3.2.6.1 MODELLE ZUR GC-CLUSTER-INSERTION

Allen GC-Clustern geht ein charakteristisches Dinukleotid voraus, welches an der Insertionsstelle der Cluster bereits gefunden und am Ende der inserierten Cluster wiederholt wird. Insbesondere V-Cluster machten deutlich, daß die terminalen Dinukleotide aus der Verdoppelung des Target-Dinukleotids hervorgehen. Zur Insertion der Cluster schlage ich daher das in Abb.36 dargestellte Modell vor.

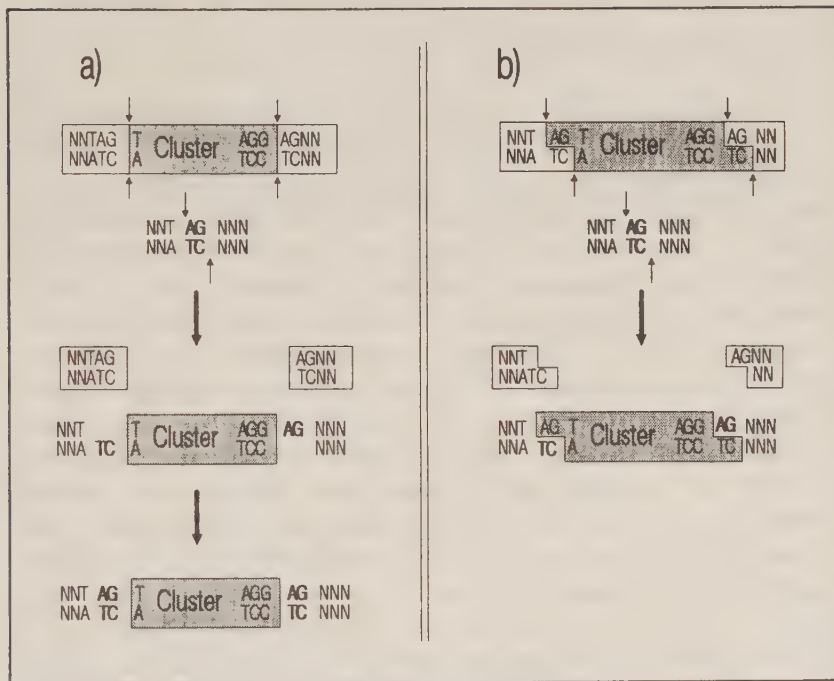


Abb.36 Modelle zu GC-Cluster-Integration

(Siehe Text)

Die Abb.36 zeigt zwei mögliche Modi am Beispiel der verbreiteten Klasse U-Cluster. Während im Modus a) die zu transponierenden Cluster durch glatte Schnitte exzidiert werden, erfolgen die Schnitte in Modus b) um zwei Nukleotide versetzt. Entsprechend verlangt Modus b) keine Verdoppelung des Targetdinukleotids. Damit ist Modus b) nur in der Lage, die Transposition von Klasse U- und Klasse G-Clustern zu beschreiben. Da er die strikte Determiniertheit der Clusterorientierung durch die Targetsequenz impliziert, ist er nicht geeignet, die Insertion von V-Clustern zu beschreiben. Dagegen beschreibt Modus a) sowohl die Situation aller drei Cluster-Klassen, als auch jene vieler transponierbarer Elemente in Prokaryonten.

3.2.6.2 TRANSPOSITIONSMODELLE IM VERGLEICH

Die einfachste Art, sich die Übertragung des GC-Clusters auf seine Targetsequenz vorzustellen, besteht in der Annahme von dsDNA-Schnitten sowohl an den beiden Rändern des Clusters (Donor-DNA), als auch in der Targetsequenz (Akzeptor-DNA). Dieser Mechanismus entspricht einer **konservativen Transposition**. Nach Ligation der Clusterenden mit den entsprechenden Enden der Targetsequenz (und Auffüllen der zurückversetzten Einzelstränge bei Modus a) in Abb.36) wäre das Transpositionereignis abgeschlossen. Ein ähnliches Modell wurde für die Transposition von Tn10 (81) vorgeschlagen. Bender und Kleckner (6) zeigten in einem eleganten Versuch, daß beide DNA-Stränge von Tn10 auf das Target Replikon übertragen werden. Dieser einfache 'schneide und übertrage'-Mechanismus impliziert dsDNA Schnitte an den Rändern der mobilen Elemente. Diese konnten bei Tn10 (81) und GC-Clustern (R. Butow, pers. Mittlg.) in vivo festgestellt werden. Ohne Ligation der Donor-DNA würde diese abgebaut. Das mobile Element (hier GC-Cluster) müßte daher von einem Replikon auf ein zweites übertragen werden. Die Ligation der Donor-DNA scheint unwahrscheinlich, da bei Erhaltung des Donor-Replikons die GC-Cluster ihr Genom mindestens so oft verlassen würden, wie sie integrieren könnten. Ihre Ausbreitung ließe sich dann nur erklären durch einen Selektionsvorteil der Cluster, bzw. durch die Replikation des exzidierten Clusters vor seiner Integration.

Eine Reihe mobiler Elemente, wie z.B. Phage Mu, IS-Elemente und ihre komplexen Transposons, sind in der Lage, in Akzeptorsequenzen zu integrieren ohne die Donor-DNA zu verlassen, wobei einige der Elemente auch in der Lage sind

Cointegrate (fusionierte Replikons) zu bilden. Der korrespondierende Mechanismus entspricht daher einer **replikativen Transposition**. Alle dazu vorgeschlagenen Modelle implizieren Einzelstrang-Schnitte an den 5' bzw. 3'Enden der mobilen Elemente. Nach Ligation des freigewordenen Endes mit der ebenfalls geöffneten Targetsequenz entsteht eine Replikationsgabel, welche die Replikation des mobilen Elements erlaubt. **Asymmetrische Modelle** beinhalten die Übertragung von zunächst nur einem Strang und anschließende Replikation an der zuerst entstehenden Replikationsgabel. Im Gegensatz dazu gehen **symmetrische Modelle** von der gleichzeitigen Übertragung der beiden Einzelstränge aus (zur Übersicht vgl. (103, 46)).

Beide replikativen Modelle sind in der Lage, sowohl einfache Insertionen als auch die Ausbildung von Cointegraten zu erklären. Degeneration von Donor-DNA, konservative Transposition und die bei Tn10 und GC-Clustern beobachteten dsDNA-Brüche an den Rändern der mobilen Elemente können jedoch nur von symmetrischen Modellen zufriedenstellend erklärt werden.

In Abb.37 ist das Shapiro-Modell (104) und eine Modifikation (87) wiedergegeben.

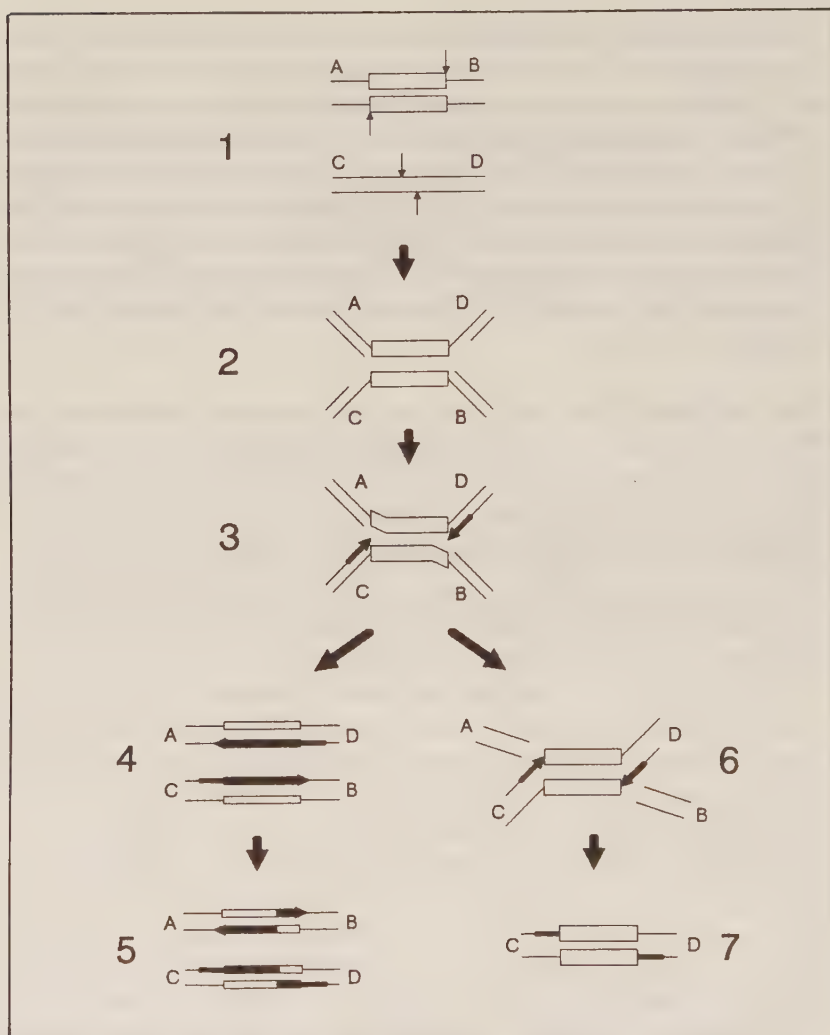


Abb.37 Symmetrische replikative Transposition

Offene Balken zeigen das ursprüngliche, dunkle Balken das neu synthetisierte Transposon. Dicke Pfeillinien deuten die Replikation an der Targetsequenz an. Die dünnen Pfeile weisen auf die Schnittstellen (nach (46) verändert; siehe Text).

Stufe 1-3: Nach den Schnitten an den 3'Enden des Transposons werden die freigebliebenen Transposonenden gleichzeitig auf ihr Target übertragen. Sie ligieren mit dem jeweils vorstehenden Ende der versetzt geschnittenen Target-DNA, wodurch zwei Replikationsgabeln gebildet werden. Das Ausmaß der daraufhin einsetzenden Replikation entscheidet über den weiteren Verlauf der Transposition.

Stufe 4-5: Das gesamte Transposon wird repliziert, wodurch ein Cointegrat entsteht. Die Transposition ist dann erst nach der Auflösung des Cointegrats beendet. Diese wird durch Rekombination der beiden Transposons an spezifischen Sequenzen (res-site) bewirkt.

Stufe 6-7: Nur die zurückversetzte Targetsequenz und eventuell ein Teil des Transposons werden repliziert. Der verdrängte DNA-Strang wird dann geschnitten, wobei die freiwerdende Donor-DNA diffundieren kann. Da nur wenige oder keine Transposonanteile repliziert werden, entspricht dieser Mechanismus zugleich konservativer Transposition: beide DNA-Stränge des Transposons werden physikalisch auf ihr Target übertragen. Die beobachteten dsDNA-Schnitte bei Tn10 und V-Clustern ließen sich auch so erklären.

Zusammengenommen weist das Gros der heute diskutierten Transpositionsmechanismen dem Donor-Replikon eines von zwei möglichen Schicksalen zu:

- (i) Das mobile Element verläßt das Donormolekül. Das Donor-Replikon wird in diesem Falle vermutlich zerstört. Die Enden können dabei durch gleichzeitige Übertragung der 3'Enden ihrer Einzelstränge (symmetrische Transposition) oder durch Übertragung des Doppelstranges (konservative Transposition) in ihr Target integrieren.
- (ii) Das mobile Element bleibt an der Donorstelle erhalten. Die Transposition folgt in diesem Falle dem Cointegrat-Modus der replikativen Modelle. Asymmetrische Modelle könnten auf das Auftreten von Cointegraten verzichten. Dabei dürfte jedoch nur ein DNA-Strang des Transposons auf das Target übertragen werden. Replikation würde dann die Transposon-DNA zum Doppelstrang ergänzen.

3.2.7 DELETION DURCH TRANSPOSITION

3.2.7.1 INTRAMOLEKULARE TRANSPOSITION

Im Folgenden wird die Situation einer intramolekularen Transposition diskutiert, wobei sich Donor- und Akzeptorstelle auf dem gleichen Replikon befinden.

Je nach Orientierung der Targetsequenz führt intramolekulare Transposition entweder zu Deletion (gleiche Orientierung) oder zu Inversion (unterschiedliche Orientierung) des dazwischenliegenden Bereiches. In Abb.37 werden beide Möglichkeiten schematisch wiedergegeben. Beide Möglichkeiten folgen dem Cointegrat-Modus, einfache Integration würde voraussichtlich zur Zerstörung des Replikons führen.

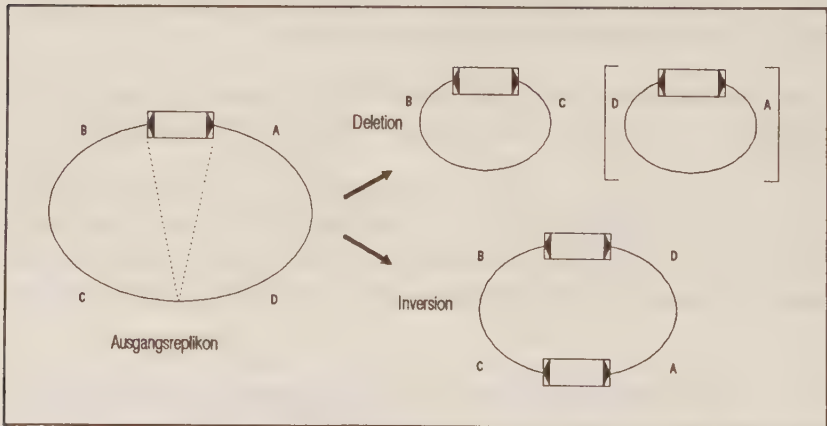


Abb.37 Intramolekulare Transposition

Das transponierbare Element (IS) ist als heller Balken wiedergegeben. Dunkle Dreiecke zeigen die Orientierung der terminalen invertierten Sequenzen an. Die beiden möglichen Schicksale des Ausgangsreplikons (links) nach der Transposition sind recht oben bzw. rechts unten dargestellt (nach (47) verändert) (siehe Text).

Beide Rearrangements wurden in IS-Elementen (IS1, IS903, IS10, IS102) häufig festgestellt, wobei Deletionen erheblich häufiger auftreten als Inversionen. An dem diesbezüglich am besten untersuchten IS903 konnte gezeigt werden, daß Deletionen etwa 100mal öfter vorkommen (86). Dies ist insofern überraschend, als wegen der statistischen Orientierung der Targetsequenz Inversionen und Deletionen mit vergleichbarer Häufigkeit zu erwarten wären. Die Gründe für die offensichtliche Bevorzugung der Deletion sind bisher nicht bekannt.

Wie in Abb.37 (oben rechts) gezeigt, sollte Deletion zu zwei reziproken Replikons führen, da nach der Transposition auch der deletierte DNA-Bereich zu einem Ringmolekül geschlossen wird. Die Replikationsfähigkeit der beiden Replikons würde dann darüber entscheiden, welche der beiden Varianten tatsächlich angetroffen wird. Neuere Untersuchungen an IS102 konnten dies bestätigen (9). Die Autoren konstruierten IS102 enthaltende Plasmide mit zwei Replikationsstart-Sequenzen (Bireplikons). Es konnte gezeigt werden, daß im Falle gleichstarker (identischer) Replikationsstarts tatsächlich beide Deletionsvarianten gleich häufig auftreten. Unterschiedlich starke Replikationsstarts resultieren in einer entsprechenden Häufigkeitsverteilung der beiden Deletionsvarianten (9).

Eine alternative Möglichkeit zur Deletionsausbildung könnte in unvollständiger Transposition bestehen, wobei nur eines der beiden Enden des mobilen Elements die Verbindung mit seiner Targetsequenz eingeht (46, 47, 9). Der deletierende Bereich (vgl. Abb.37, oben rechts) würde dann nicht zyklisieren und daher abgebaut.

3.2.7.2 MITOCHONDRIALE DELETIONEN (MIT-DELA, RHO⁻)

In der vorliegenden Arbeit konnte gezeigt werden, daß GC-Cluster zugleich die Targetsequenz weiterer GC-Cluster Integration beinhalten (vgl. Abschn. 3.2.3.1.4). Diese Targets werden von U1a-Clustern gelegentlich, von U2a-Clustern häufig und von U3-Clustern fast ausschließlich gewählt.

Die Deletionsklasse A, deren häufige Deletion zwischen zwei U2a-Clustern auftritt (vgl. Abschn. 3.1.4.2), entspricht damit den von IS-Elementen verursachten Deletionen. Sie könnte aus intramolekularer Transposition des Clusters III (a15b)

an die TAG-Targetsequenz am 5'Anfang des Clusters V (OXI3-Leader) hervorgehen.

Ähnliches gilt für die Exzision von rho-Genomen, welche vorwiegend an GC-Clustern beobachtet wurde (4, 38). Beim Vergleich mitochondrialer DNA verschiedener rho-Stämme konnte ich darüberhinaus auch Deletionen finden, welche vom Ende eines GC-Clusters bis zu einer (clusterfreien) TAG-Targetsequenz reichten (Daten nicht gezeigt). Diese Deletionen lassen sich nicht durch die Rekombination homologer Bereiche erklären, wie dies bisher für die Exzision an GC-Clustern exzisierten rho- diskutiert wurde (4). Alle diese Beobachtungen lassen sich jedoch durch intrachromosomale Transposition von GC-Clustern erklären, wobei die Targetsequenz der Transposition selbst ein GC-Cluster darstellen kann.

3.3 ZUSAMMENFASSUNG

3.3.1 SEQUENZ-SPEZIFISCHE DELETIONEN

Genetische Untersuchungen ließen erkennen, daß mit Deletionen im OXI3-Gen mit besonderer Häufigkeit auftreten. Durch die Feinkartierung von OXI3-Deletionsmutanten konnte gezeigt werden, daß vorwiegend zwei Gruppen von Mutanten mit vermutlich sequenz-spezifischen Deletionen angetroffen werden (19, Sang Ho Kim, pers. Mittlg.). Diese wurden mit Klasse A (mit-delA) bzw. Klasse B (mit-delB) bezeichnet. Alle Mutanten wurden durch Mn²⁺-Mutagenese aus dem wt 777-3A abgeleitet (vgl. Abschn. 3.1.1).

Durch die Sequenzierung von sieben Deletionsmutanten konnte ich die Sequenzspezifität der Deletionen bestätigen vgl. Abschn. 3.1.4.1 bzw. 3.1.5.1). In jeder der beiden Klassen waren die Mutanten aus identischen Deletionsereignissen hervorgegangen.

Die mit-delB-Deletion (Klasse B) umspannt 8221 bp. Die Deletion beginnt am Ende des Introns a11 mit der ersten Base des Exons A2, sie endet im Intron a15b (vgl. Abschn. 3.1.5.1). Das Ende des deletierten Bereichs von Intron a15b besitzt auffällige Homologie (14 von 19 bp) mit dem Ende von Exon A1. Die Klasse B-Deletion erinnert daher an das Spleißen von Intron a11 (vgl. Abschn. 3.1.5.2.2). Introndeletionen wurden in unterschiedlichen Organismen häufig beobachtet. Anders als dort deletieren bei der Klasse B jedoch Sequenzen, welche den Exons entsprechen. Hypothesen, welche bisher die Deletion von Introns zu erklären versuchten (40), sind ohne weitgehende Modifikationen nicht in der Lage, die Klasse B-Deletion zu erklären (vgl. Abschn. 3.1.5.2.3). Mögliche Mechanismen der Klasse B Deletion wurden in dieser Arbeit diskutiert: sie umfassen sowohl Heteroduplexstrukturen aus fehlgespleißter RNA und mtDNA (vgl. Abschn. 3.1.5.2.5), als auch die direkte Exzision der DNA ohne Beteiligung von RNA (vgl. Abschn. 3.1.5.2.6).

Die Deletionen der mit-delA-Mutanten (Klasse A) umfaßt 11 347 Nukleotide: sie beginnt bei einem GC-Cluster im OXI3-Leader und endet bei einem der optionalen GC-Cluster im Intron a15b. Beginnend mit dem Anfang der beiden GC-Cluster sind 31 Nukleotide an den beiden Deletionsenden identisch. Innerhalb

dieses homologen Bereiches finden offensichtlich jene Rekombinationsereignisse statt, welche zur Klasse A-Deletion führen. In den mit-delA-Mutanten bleibt eine der beiden homologen Sequenzen zurück (vgl. Abschn. 3.1.4.1).

Im Falle der mit-delB-Mutanten konnten die Deletionsendpunkte durch den Vergleich der Nahtstellensequenz mit der bereits publizierten Sequenz des wt KL14-4A ermittelt werden. Bei den mit-delA-Mutanten war dies nicht möglich. Das GC-Cluster am linken mit-delA-Deletionsende fiel in eine bisher nicht bestimmte Sequenzlücke im OXI3-Leader. Das rechte mit-delA-Deletionsende betrifft ein GC-Cluster, welches in der Sequenz des bereits sequenzierten Stammes KL14-4A nicht auftritt. Ich war daher gezwungen, die beiden an der Deletion beteiligten Bereiche im wt 777-3A zu sequenzieren.

Im Rahmen dieser Sequenzierungsarbeiten stellte ich fest, daß 777-3A sowohl im Intron aI5b als auch im Bereich des OXI3-Leaders weitere optionale GC-Cluster besitzt. Dies führte zu der Idee, daß es sich bei GC-Clustern um mobile Elemente handeln könnte, deren Transposition zu den spezifischen Deletionen der Klasse A führen könnte.

Ich erhoffte mir, durch die Sequenzierung der optionalen GC-Cluster (GC⁺-Sequenz) und dem Sequenzvergleich mit den GC-Cluster freien Allelen (GC⁻-Sequenz) anderer Hefestämme, Aussagen über die Targetsequenzen der GC-Cluster machen zu können. Daher sequenzierte ich im wt 777-3A sowohl das Intron aI5b, als auch den OXI3-Leader und einen 500 bp langen Bereich, der dem Leader vorausgeht.

3.3.2 SEQUENZIERUNG DES WT 777-3A

Die Sequenzanalyse ergab, daß das **Intron aI5b** vier GC-Cluster enthält, wobei drei der Cluster optional sind. Das Intron besitzt einen langen offenen Leserahmen (ORF), welcher vermutlich eine RNA-Maturase codiert (50). Das in beiden Stämmen auftretende GC-Cluster (Cluster I) und zwei der drei optionalen GC-Cluster (Cluster II und III) betreffen diesen ORF, wobei letzteres eine -1-Verschiebung und damit eine vorzeitige Terminierung des Leserahmens bewirkt (vgl. Abschn. 3.1.2.2).

Durch die Sequenzierung des **OXI3-Leaders** (vgl. Abschn. 3.1.3.1) konnte ich eine 105 bp lange Sequenzlücke schließen. Ein GC-Cluster (Cluster V), welches zugleich das linke Deletionsende der Klasse A-Mutanten bildet, betrifft diese Lücke. Ein 500 bp langer Bereich, der dem Leader vorausgeht, wurde ebenfalls von mir sequenziert. 108 bp vor dem OXI3-Promotor (Pos. -622) konnte ich in 777-3A ein weiteres GC-Cluster bestimmen (Cluster VI). Dieses Cluster ist optional, da es in der Sequenz des wt KL14-4A fehlt (vgl. Abschn. 3.1.3.2).

3.3.3 TAGET- UND INSERTIOSSEQUENZEN DER GC-CLUSTER

Ich hatte nun vier bisher nicht bekannte **optionale GC-Cluster** sequenziert. Aus dem Vergleich dieser und drei weiterer, aus der Literatur bekannter, optionaler GC-Cluster mit den entsprechenden GC-Sequenzen anderer Hefestämme, konnte ich erstmals erkennen, daß die Cluster in die Konsensussequenz TAG integriert sind. Es war bereits bekannt, daß dieses TAG auch am 5'Anfang vieler nicht-optionaler GC-Cluster konserviert ist (107). Hieraus leitete ich die **Targetsite-Hypothese** ab, welche die GC-Clustersequenzen in Targetsequenz (TAG) und eigentliche Insertionssequenz differenziert (vgl. Abschn. 3.2.1.2).

Auf Grund dieser ermutigenden Ergebnisse, suchte ich, mittels Restriktionsanalyse und Southernblot-Hybridisierung 15 verschiedener Hefestämme, nach weiteren optionalen GC-Clustern. Zusätzlich zu den vier o.g. Clustern gelang es mir, vier weitere optionale GC-Cluster zu sequenzieren (vgl. Abschn. 3.2.3.1.5).

Heute sind über 90% des mt Genoms der Hefe sequenziert (zur Übersicht vgl. (130, 131)). Die einzelnen Bereiche des Genoms wurden von verschiedenen Arbeitsgruppen in unterschiedlichen Hefestämmen bestimmt. Über 10% des Genoms wurden daher in wenigstens zwei Hefestämmen sequenziert. Aus dem Vergleich der überlappend sequenzierten Bereiche erhoffte ich mir, weitere optionale GC-Cluster zu finden. Daher durchsuchte ich alle verfügbaren publizierten mt Sequenzen nach mehrfach sequenzierten Bereichen. Weitere, noch nicht veröffentlichte Sequenzdaten, wurden mir von anderen Arbeitsgruppen zur Verfügung gestellt. In den in unterschiedlichen Stämmen sequenzierten Bereichen

konnte ich weitere optionale GC-Cluster finden. Damit standen mir insgesamt 19 optionale GC-Cluster und deren GC-Sequenzen zur Verfügung.

Der GC+/GC--Sequenzvergleich bestätigte die Targetsites-Hypothese. An Stelle der GC-Cluster konnte ich in den GC-Sequenzen meist das Trinukleotid TAG finden. Da der Großteil der Cluster, welche abweichende Targetsequenzen aufwiesen, sich auch in der Primärsequenz von den übrigen GC-Clustern unterschied, lag die Vermutung nahe, daß verschiedene Gruppen von GC-Clustern verschiedene Targetsequenzen benutzen.

Zu diesem Zeitpunkt meiner Arbeit waren die GC-Cluster noch nicht zentral erfaßt. Ich durchsuchte daher alle mir zur Verfügung stehenden Sequenzdaten nach GC-Clustern, gab deren Sequenzen in einen Microcomputer ein und entwickelte die zur weiteren Analyse benötigte Software (vgl. Abschn. 2.2.2).

Auf Grund ihrer Primärsequenzen konnte ich eine erste Gruppierung der Cluster vornehmen. Gleichzeitig und unabhängig wurden von Zamaroczy & Bernardi vergleichende Untersuchungen an GC-Clustern durchgeführt (129). Meine Analyse bestätigten die dort erzielten Resultate im Wesentlichen (vgl. Abschn. 3.2.4.1).

Der Vergleich der GC+/GC--Sequenzpaare ermöglichte mir erstmals die Cluster sowohl nach dem Kriterium ihrer optionalen Bestandteile, welche die vermuteten Insertionssequenzen darstellen, als auch nach dem Kriterium ihrer Targetsequenzen zu definieren. Hierdurch konnte ich eine **neue Klassifizierung der Cluster** vornehmen. Diese können nach der von mir entwickelten Systematik in die Klassen U, G, V und die imperfekte Klasse I eingeteilt werden (vgl. Abschn. 3.2.2.3 und 3.2.4). Die Cluster aller drei perfekten Klassen besitzen klassen-spezifische Targetsequenzen.

Die Klasse U, welche die überwiegende Mehrzahl aller GC-Cluster enthält, ist in weitere, in sich homogene Familien und Unterfamilien untergliedert (U1a, U1b; U2a, U2b, U2c; U3; U4; vgl. Abschn. 3.2.3.1.1). Alle U-Cluster enden mit der Sequenz AGGAG.

Ich konnte erstmals zeigen, daß U-Cluster im Anschluß an das Ende ihrer kurzen Targetsequenz TAG (bzw. AAGGAG bei U3-Clustern) inserieren. Die Klasse U-Targetsequenz wird durch die Integration der Cluster somit nicht zerstört.

Vermutlich durch die sukzessive Benutzung der selben Targetsequenz sind U-Cluster in der Lage, **komplexe Tandemaggregate** aus mehreren GC-Clustern zu bilden. Diese GC-Clusteraggregate wurden von mir erstmals systematisch analysiert (vgl. Abschn. 3.2.3.1.3 bis 3.2.3.1.5).

Die Analyse der Tandemaggregate bestätigte meine Differenzierung der GC-Clustersequenzen in Targetsequenz und Insertionssequenz. Die Targetsequenz geht jeweils nur dem ersten Cluster eines Aggregats voraus. Bei den weiter hinten liegenden Clustern, schließt deren Insertionssequenz unmittelbar an das AGGAG-Ende des vorangehenden Cluster an (vgl. Abschn. 3.2.3.1.3).

Durch die Sequenzierung von GC-Clustern, welche **optionale Bestandteile von Clusteraggregaten** bilden, konnte ich zeigen, daß die Targetsequenz der U3-Cluster (AAGGAG) von jener der übrigen U-Cluster (TAG) abweicht. Die U3-Targetsequenz ist mit dem 3'Ende der U-Cluster identisch. Es ist von daher verständlich, daß der Großteil aller U3-Cluster in bereits vorhandenen GC-Clustern inseriert ist und daher in Tandemaggregaten angetroffen wird.

Die Targetsequenz der **Klasse G-Cluster** ist erheblich komplexer. G-Cluster werden ausschließlich in den homologen Sequenzen der sog. Replikationsstart-Sequenzen gefunden. Fünf der acht bekannten Replikationsstart-Sequenzen (ORI1 bis ORI8) besitzen ein G-Cluster an einer spezifischen Position. Nur eines der Cluster ist optional. Zur Bestimmung der Klasse G-Targetsequenz konnte ich auch jene Replikationsstart-Sequenzen heranziehen, welche kein G-Cluster besitzen. Damit war ich in der Lage auch die Target- und Insertionssequenz der G-Cluster erstmals zu bestimmen (vgl. Abschn. 3.2.3.2).

Die GC-Cluster der **Klasse V** weichen erheblich von den übrigen Clustern ab. Ihre Targetsequenz ist 20 bp lang (vgl. Abschn. 3.2.3.3). Auf die Targetsequenz dieser Cluster wurde in verschiedenen Arbeiten der Arbeitsgruppe R. Butow bereits hingewiesen (19, 56).

3.3.4 INSERTIONSMECHANISMUS DER GC-CLUSTER

Durch den Vergleich der GC+/GC--Sequenzen konnte ich die exakten Endpunkte der GC-Cluster bestimmen. Alle drei GC-Cluster-Klassen (U, G, V) beginnen und

enden mit einem charakteristischen Dinukleotid (AG, GT, AA). Dieses Dinukleotid ist an der Insertionsstelle der Cluster bereits vorhanden. Es ist daher wahrscheinlich, daß die Dinukleotide an den Enden der Cluster durch die Verdoppelung des Dinukleotids an der Insertionsstelle entstanden sind. Dies erinnert an die Transposition prokaryontischer IS-Elemente, deren Integration stets durch die Verdoppelung von 3 bis 12 bp ihrer Targetsequenz begleitet wird. Diese Analogie erlaubte mir, **Modelle zur Transposition von GC-Clustern** vorzuschlagen (vgl. Abschn. 3.2.6).

Während intermolekulare Transposition in der Ausbreitung der Cluster resultiert, würde intramolekulare Transposition zur **Deletion** des dazwischenliegenden DNA-Bereichs führen (vgl. Abschn. 3.2.7.1). Ich konnte in dieser Arbeit zeigen, daß die Ränder der U-Cluster zugleich die Targetsequenzen weiterer U-Cluster-Integration darstellen. Diese Targetsequenzen werden von einigen Klasse U-Familien bevorzugt genutzt. Intramolekulare Transposition eines U-Clusters an die Targetsequenz eines zweiten U-Clusters würde zum Verlust des dazwischenliegenden Bereichs führen, wobei eines der Cluster in der Deletionsmutante zurückbliebe (vgl. Abschn. 3.2.7.2). Die häufigen Deletionen der o.g. Deletionsklasse A lassen sich so erklären.

No.	Type	opt.	location	strand	strain	ref.	No.	Type	opt.	location	strand	strain	ref.
1	U3	o	21S+5	tr	777-3A	a)	48	U1a		15S-400	?	C982-19d	43
2	I1		21S+35	tr	777-3A	a)	49	U3		15S-455	tr	MH41-7B	108
3	U4		21S+100	tr	777-3A	a)	50	U2c		15S-440	tr	MH41-7B	108
1-			21S+5	tr	J18-8C	109	51	U3		15S-375	tr	MH41-7B	108
2'	I1		21S+35	tr	JL8-8C	109	52	U2a		15S-320	tr	MH41-7B	108
3'	U4		21S+100	tr	JL8-8C	109	53	U1a	o	15S 450	nt	MH41-7B	108
4	U3		thr2-190	tr	D273-10B	7						MH41-7B-P2	b)
5	U2c		leu-950	tr	D273-10B	8						D273-10B	66
6	U2a		leu-910	tr	D273-10B	8	53-			15S 450	nt	D243-2B-R1	b)
7	U2a		leu-340	nt	D273-10B	8						FF1210-6c	108
8	U2a		leu-190	nt	D273-10B	8	54	U4		15S+40	nt	D243-2B-R1	71
9	U2a		gln+150	tr	D273-10B	8	55	U1a		15S+95	nt	D243-2B-R1	71
10	U2c		lys-335	tr	D273-10B	16	56	U2c		ORF3 140	tr	D243-2B-R1	71
11	U3		asp+325	tr	D273-10B	16	57	U2a	o	ORF3 170	tr	D243-2B-R1	71
12	I3		asp+360	tr	D273-10B	16	56'	U2c		ORF3 140	tr	D8-25A	d.A.
13	U1a		ser2-150	nt	D273-10B	16	57-			ORF3 170	tr	D8-25A	d.A.
14	U1b	o	ile+130	nt	SM202	b)	58	U3	o	ORF3+160	tr	D243-2B-R1	71
14-			ile+130	nt	SM201	b)	59	I		ORF3+205	tr	D243-2B-R1	71
15	U3		ile+275	tr	SM202	b)	58-			ORF3+160	tr	D8-25A	d.A.
16	U2a		ile+315	tr	SM202	b)	59'	I		ORF3+205	tr	D8-25A	d.A.
17	U1a	o	asn-180	tr	D273-10B	24	60	U2a	o	trp-235	tr	D243-2B-R1	71
17-			asn-180	tr	SM202	b)	60-			trp-235	tr	TR3-15A	b)
18	U2c		asn-70	tr	D273-10B	24	61	U3	o	trp+80	tr	D243-2B-R1	71
19	I4		met+70	tr	D273-10B	24	62	U3		trp+125	tr	D243-2B-R1	71
20	U2c		met+85	tr	D273-10B	24	61-			trp+80	tr	TR3-15A	b)
21	U3		met+150	tr	D273-10B	24	62'	U3		trp+125	tr	TR3-15A	b)
22	U2a		met+180	tr	D273-10B	24	63	U2a	o	trp+385	tr	D243-2B-R1	71
23	I1		OXI1-320	nt	D273-10B	24	63-			trp+385	tr	777-3A	b)
24	U3		OXI1-290	nt	D273-10B	24	64	U4	o	trp+515	nt	D243-2B-R1	71
25	U1a		ORF1 900	tr	D273-10B	24	65	U1a	o	trp+565	nt	D243-2B-R1	71
26	U1a		OXI2+145	nt	D273-10B	78	64-, 65-			trp+515	nt	D8-25A	d.A.
27	U1b		ORF2 240	tr	D273-10B	78	66	U1a		ori8-170	tr	D243-2B-R1	34
28	U1b		ORF2 920	tr	D273-10B	78	67	U3	o	ori8 175	tr	D243-2B-R1	34
29	U1b		ORF2+195	tr	D273-10B	78	68	U2a	o	ori8 215	tr	D243-2B-R1	34
30	U2a		ori5-930	tr	D243-2B-R1	43	67-, 68-			ori8 175	tr	C982-19d	34
31	U3		ori5-830	tr	D243-2B-R1	43	69	6		ori8 300	tr	C982-19d	34
32	I4		ori5-630	tr	D243-2B-R1	43				ori8 400	tr	D243-2B-R1	
33	U1a		ori5-540	tr	D243-2B-R1	43	70	U2b	o	ori8 530	tr	D243-2B-R1	34
34	U3		ori5-415	nt	D243-2B-R1	43	70-			ori8 530	tr	C982-19d/	34
35	U3		ori5+250	nt	KL14-4A	14						KL14-4A	
36	U3		fmet+60	tr	MH41-7B	80	71	U1a	o	OXI3-620	nt	777-3A	d.A.
37	V	o	9S-60	nt	MH41-7B	80	71-			OXI3-620	nt	KL14-4A	89
37-			9S-60	nt	D273-10B	80	72	U2a		OXI3-200	nt	777-3A	d.A.
38	U2a		9S 300	nt	MH41-7B	80	73	U2a		a12 40	tr	D273-10B	17
39	U1a		9S+115	tr	MH41-7B	80	74	Iv		a13 175	nt	D273-10B	17
40	U1a		ori1-1050	nt	C982-19d	34	75	I		a13 275	nt	D273-10B	17
41	U1a	o	ori1-430	tr	KL14-4A	34	76	U1a		a15b 440	tr	777-3A	d.A.
41-			ori1-430	tr	C982-19d	34						KL14-4A	50
42	U1a	o	ori1-210	tr	D243-2B-R1	34	77	U2a	o	a15b 850	tr	777-3A	d.A.
43	U4		ori1-160	tr	D243-2B-R1	34	77-			a15b 850	tr	KL14-4A	50
42-, 43-			ori1-210	tr	KL14-4A/	34	78	U2a	o	a15b 1155	nt	777-3A	d.A.
					C982-19d		78-			a15b 1155	nt	KL14-4A	50
44	U3	o	ori1 195	tr	C982-19d	34	79	U1a	o	a15b 1470	nt	777-3A	d.A.
45	U2b		ori1 225	tr	C982-19d	34	79-			a15b 1470	nt	KL14-4A	50
46	I?		ori1 235	tr	C982-19d	34	80	U2a		OLI2-355	nt	J69-1B	85
44-			ori1 195	tr	KL14-4A	35	81	U3		OLI2-110	tr	J69-1B	85
45'	U2b	o	ori1 225	tr	KL14-4A	35	81'	I3		OLI2-110	tr	D273-10B	67
45'-			ori1 225	tr	D243-2B-R1	35	82	U1a		ORF4 265	nt	D273-10B	102
47	6	o	ori1 300	tr	KL14-4A	35	83	U2b		ORF4 535	nt	D273-10B	102
			ori1 580	tr	C982-19d		84	U1a		ORF4 1010	nt	D273-10B	102
47-			ori1 300	tr	D243-2B-R1	24	85	U1a		ori7-420	nt	C982-19d	43

No.	Type	opt.	location	strand	strain	ref.	No.	Type	opt.	location	strand	strain	ref.
86	U2a		ori7-40	nt	D273-10B	67					nt	D243-2B-R1	128
				nt	C982-19d	128	114	U1a		ori6+60	tr	D273-10B	84
87	G		ori7 25	nt	D273-10B	67					tr	D243-2B-R1	128
				nt	C982-19d	24;84	115	U2c		ori6+260	tr	D273-10B	84
88	I1b		ori7+235	nt	C982-19d	84	116	U3		ori6+325	tr	D273-10B	84
89	U1a		ORF5+70	tr	C982-19d	84	117	U2a		ori6+360	tr	D273-10B	84
90	U3		ORF5+200	tr	C982-19d	84	118	U3		ori6+600	tr	D273-10B	84
91	U3		ORF5+275	tr	C982-19d	84	119	U2c		OLI1-710	nt	D273-10B	118
92	I3		ORF5+670	tr	C982-19d	84	120	U1a		OLI1+570	tr	D273-10B	118
93	I		ORF5+710	tr	C982-19d	84	121	V		OLI1+770	nt	D273-10B	118
94	U3		ori2+900	nt	C982-19d	43	122	U3		ser1-300	nt	992	121
95	U1a		ori2+1000	tr	C982-19d	43	123	U3	o	ser1-245	nt	992	121
96	U3		ori2+1300	nt	C982-19d	43	122	U3		ser1-300	nt	D273-10B	118
97	U2b		ori2+1700	nt	C982-19d	43	123			ser1-245	nt	D273-10B	118
98	U1a		ori2+1800	nt	C982-19d	43	124	U1a		ser1-100	nt	D273-10B	118
99	I3		ori2+1900	nt	C982-19d	43	125	U2a		ser1+155	nt	D273-10B	56
100	I1a		ori2+2000	nt	C982-19d	43	126	U2c		ser1+190	nt	D273-10B	56
101	U1a		glu-435	nt	C982-19d	43	127	U3		ser1+205	nt	D273-10B	56
				nt	777-3A	82	128	V		VAR1 200	nt	A10	55
102	U3		glu-275	nt	777-3A	82						D273-10B	118
103	U3		glu-230	nt	777-3A	82						J ^f , J ^s , 50SS	56
104	U1a		glu+365	nt	777-3A	30						NP1-1B, D6	56
105	U2a		bI3 1140	nt	777-3A	53	129	V	o	VAR1 440	tr	J ^f	56
106	U3	o	bI3 1180	nt	777-3A	53						D6, 50SS	56
107	U3		bI3 1250	nt	777-3A	53	129			VAR1 440	tr	A10	55
105'	U2a		bI3 1140	nt	D8-25A	d.A.						D273-10B	118
106			bI3 1180	nt	D8-25A	d.A.						NP1-1B, J ^s	65
107'	U3		bI3 1250	nt	D8-25A	d.A.	130	U1a		VAR1+15	nt	D273-10B	118
108	U1a		bI5 540	nt	D273-10B	84	131	I3		ori3-240	tr	KL14-4A	13
109	U1a		C08+330	nt	D273-10B	84	132	U2b		ori4 140	tr	D243-2B-R1	128
110	U3		C08+505	nt	D273-10B	84	133	G		ori4+40	tr	D243-2B-R1	128
111	I3		C08+830	tr	D273-10B	84	134	U3		21S-725	tr	JL8-8C	109
112	G		ori6 25	nt	D243-2B-R1	128	135	U2c		21S-215	nt	JL8-8C	109
113	U2b		ori6 130	nt	D273-10B	84	136	U3		21S-200	nt	JL8-8C	109
							137	I	o	21S 2530	tr	JL8-8C	31
							137			21S 2530	tr	JL9-8A/	31
												JL16-11D	

Verzeichniss aller erwähnten GC-Cluster

- Spalte 1: Nummer des Clusters
 Indizes: n' = variante Form des Clusters n
 n = GC-Sequenz des Clusters n
- Spalte 2: Klassifikation des Clusters
- Spalte 3: Localisation des Clusters
- Spalte 4: o = optionales Cluster
- Spalte 5: Das angegebene Cluster gibt die Sequenz des transkribierten (tr) bzw. nicht-transkribierten (nt) DNA-Stranges wieder
- Spalte 6: Hefestämme, welche das Cluster enthalten
- Spalte 7: Quelle (vgl. Anhang B)
 a) = B. Dujon pers. Mittlg.
 b) = R. Martin pers. Mittlg.
 d.A = diese Arbeit

LITERATUR

- (1) Albertini, A., Hofer, M., Calos, M. P. & Miller, J. (1982) On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions, *Cell* 29, 319-328
- (2) Anderson, S., Bankier, A. T., Barrell, B. G., De Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome, *Nature (London)* 290, 457-465
- (3) Ansorge, W. & Maeyer, L. (1980) Thermally stabilized very thin (0.02-0.3mm) polyacrylamide gels for electrophoresis, *Journal of Chromatography* 202, 45-53
- (4) Baldacci, G., de Zamoroczy, M. & Bernardi, G. (1980) Excision sites in the GC clusters of the mitochondrial genome of yeast, *FEBS Letters* 114, 234-236
- (5) Belcour, L. & Vierny, C. (1986) Variable splicing sites of a mitochondrial intron: Relationship to the senescence process in *Podospora*, *EMBO J.* 5, 609-614
- (6) Bender, J. & Kleckner, N. (1986) Genetic evidence that Tn10 transposes by a nonreplicative mechanism, *Cell* 45, 801-815
- (7) Berlani R. E., Pentella, C., Macino, G. & Tzagoloff, A. (1980) Assembly of the mitochondrial membrane system: Isolation of mitochondrial transfer ribonucleic acid mutants and characterization of transfer ribonucleic acid genes of *Saccharomyces cerevisiae*, *J. Bacteriol.* 141, 1086-1097
- (8) Berlani, R. E., Bonitz, S. G., Coruzzi, G., Nobrega, M. & Tzagoloff, A. (1980) Transfer RNA genes in the cap-oxi1 region of yeast mitochondrial DNA, *Nucl. Acids Res.* 8, 5017-5030
- (9) Bernardi, F. & Bernardi, A. (1986) Intramolecular transposition of IS102, *Gene* 42, 11-19
- (10) Bernardi, G. (1979) The petite mutation in yeast, *TIBS* 4, 197-201
- (11) Bibb, M. J., Etten van, R. A., Wright, C. T., Walberg, M. W. & Clayton, D. A. (1981) Sequence and gene organization of mouse mitochondrial DNA, *Cell* 26, 167-180

- (12) Birnboim, H. C. & Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA, *Nucl. Acids Res.* 7, 1513-1523
- (13) Blanc, H. & Dujon, B. (1980) Replicator regions of the yeast mitochondrial DNA responsible for suppressiveness, *Proc. Natl. Acad. Sci. U.S.A.* 77, 3942-3946
- (14) Blanc, H. (1984) Two modules from the hypersuppressive rho⁻ mitochondrial DNA are required for plasmid replication in yeast, *Gene* 30, 47-61
- (15) Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. (1985) Ty elements transpose through an RNA intermediate, *Cell* 40, 491-500
- (16) Bonitz, S. G. & Tzagoloff, A. (1980) Assembly of the mitochondrial membrane system: Sequences of yeast mitochondrial tRNA genes, *J. Biol. Chem.* 255, 9075-9081
- (17) Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagoloff, A. & Macino, G. (1980) Assembly of the mitochondrial membrane system: Structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase, *J. Biol. Chem.* 255, 11927-11941
- (18) Bruckner, H. (1983) Die OXI3-Region der mtDNA der Hefe: Genetische Kartierung und Nachweis Sequenz-spezifischer Deletionen, Dissertation, LMU München
- (19) Butow, R. A., Perlman, P. S. & Grossman, L. I. (1985) The unusual var1 Gene of yeast mitochondrial DNA, *Science* 228, 1496-1501
- (20) Cech, T. R. & Bass, B. L. (1986) Biological catalysis by RNA, *Annu. Rev. Biochem.* 55, 599-629
- (21) Cech, T. R., Zaugg, A. J. & Grabowski, P. J. (1981) In vitro splicing of the ribosomal precursor of Tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence, *Cell* 27, 487-496
- (22) Claasen, S. (1985) Häufige, exakte Deletion von Introns der mitochondrialen DNA von *Saccharomyces cerevisiae*, Diplomarbeit, LMU München
- (23) Colin, Y., Baldacci, G. & Bernardi, G. (1985) A new putative gene in the mitochondrial genome of *Saccharomyces cerevisiae*, *Gene* 36, 1-13

- (24) Coruzzi, G., Bonitz, S. G., Thalenfeld, B. E. & Tzagoloff, A. (1981) Assembly of the mitochondrial membrane system: Analysis of the nucleotide sequence and transcripts in the *oxi1* region of yeast mitochondrial DNA, *J. Biol. Chem.* 256, 12780-12787
- (25) Cosson, J. & Tzagoloff, A. (1979) Sequence homologies of G+C rich regions of mitochondrial DNA of *Saccharomyces cer.*, *J. Biol. Chem.* 254, 42-43
- (26) Craig, N. L. (1985) Site-specific inversion: Enhancers, recombination proteins, and mechanism, *Cell* 41, 649-650
- (27) Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. (1982) Making ends meet: A model for RNA splicing in fungal mitochondria, *Nature (London)* 300, 719-724
- (28) De la Salle, H., Jacq, C. & Slonimski, P. P. (1982) Critical sequences within mitochondrial introns: Pleiotropic mRNA maturase and cis-dominant signals of the box intron controlling reductase and oxidase, *Cell* 28, 721-732
- (29) Dinter-Gottlieb, G. (1986) Viroids and virusoids are related to group I introns, *Proc. Natl. Acad. Sci. U.S.A.* 83, 6250-6254
- (30) Dobres, M., Gerbl-Rieger, S., Schmelzer, C., Mueller, M. W. & Schweyen, R. J. (1985) Deletions in the *cob* gene of yeast mtDNA and their phenotypic effect, *Curr. Genet.* 10, 283-290
- (31) Dujon, B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the omega and rib-1 loci, *Cell* 20, 185-197
- (32) Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., de Riel, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980) The structure and evolution of the human α -globin gene family, *Cell* 21, 653-668
- (33) Fangman, W. L. & Dujon, B. (1984) Yeast mitochondrial genomes consisting of only A-T bases replicate and exhibit suppressiveness, *Proc. Natl. Acad. Sci. U.S.A.* 81, 7156-7160
- (34) Faugeron-Fonty, G. & Goyon, C. (1985) Polymorphic variations in the *ori* sequences from the mitochondrial genomes of different wild-type yeast strains, *Curr. Genet.* 10, 269-282

- (35) Faugeron-Fonty, G., Le Van Kim, C., de Zamaroczy, M., Goursot, R. & Bernardi, G. (1984) A comparative study of the ori sequences from the mitochondrial genome of twenty wild-type yeast strains, *Gene* 32, 459-473
- (36) Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Merregaert, J., Min Jou, W., Raeymakers, A., Volckaert, G., Ysebaert, M., Kerckhove van de, J., Nolf, F. & Montagu van, M. (1975) A protein gene of bacteriophage MS2, *Nature (London)* 256, 273-278
- (37) Furano, A. V., Somerville, C. C., Tschlis, P. N. & D'Ambrosio, E. (1986) Target sites for the transposition of rat long interspersed repeated DNA elements (LINEs) are not random, *Nucl. Acids Res.* 14, 3717
- (38) Gaillard, C., Strauss, F. & Bernardi, G. (1980) Excision sequences in the mitochondrial genome of yeast, *Nature (London)* 283, 218-220
- (39) Garfinkel, D. J., Boeke, J. D. & Fink, G. R. (1985) Ty element transposition: Reverse transcriptase and virus-like particles, *Cell* 42, 507-517
- (40) Gargouri, A., Lazowska, J. & Slonimski, P. P. (1983) DNA-splicing of introns in the gene: General way of reverting intron mutation, in *Mitochondria 1983: (Schweyen R.J., Wolf K., Kaudewitz F. eds) Walter de Gruyter - Berlin - New York*, 259-268
- (41) Garriga, G., Lambowitz, A. M., Inoue, T. & Cech, T. R. (1986) Mechanism of recognition of the 5'splice site in self-splicing group I introns, *Nature (London)* 322, 86
- (42) Gough, J. & Murray, N. E. (1983) Sequence diversity among related genes for recognition of specific targets in DNA molecules, *J. Mol. Biol.* 166, 1-19
- (43) Goursot, R., Mangin, M. & Bernardi, G. (1982) Surrogate origins of replication in the mitochondrial genomes of ori^s petite mutants of yeast, *EMBO J.* 1, 705-711
- (44) Green, M. R. (1986) Pre-mRNA splicing, *Annu. Rev. Genet.* 20, 671-708
- (45) Greer, C. L., Peebles, C. L., Gegenheimer, J. & Abelson, J. (1983) Mechanism of action of a yeast RNA ligase in tRNA splicing, *Cell* 32, 537-546
- (46) Grindley, N. D. F. & Reed, R. R. (1985) Transposable recombination in prokaryotes, *Annu. Rev. Biochem.* 54, 863-896

- (47) Grindley, N. D. F. & Sherratt, D. J. (1979) Sequence analysis at IS1 insertion sites: Models for transposition, Cold Spring Harbor Symp. Quant. Biol. 43, 1253-1261
- (48) Grundstein, M. & Hogness, D. (1975) Colony hybridization: A method for selection of cloned DNAs that contain a specific gene, Proc. Natl. Acad. Sci. U.S.A. 72, 3961
- (49) Haid, A., Schweyen, R. J., Bechmann, H., Kaudewitz, F., Solioz, M. & Schatz, G. (1979) The mitochondrial COB region in yeast codes for apocytochrome b and is mosaic, Eur. J. Biochem. 94, 451-464
- (50) Hensgens, L. A. M., Bonen, L., de Haan, M., van der Horst, G. & Grivell, L. A. (1983) Two intron sequences in yeast mitochondrial *cox1* gene: Homology among urf-containing introns and strain-dependent variation in flanking exons, Cell 32, 379-389
- (51) Hirochika, H., Takatsuji, H., Ubasawa, A. & Ikeda, J. (1985) Site-specific deletion in cauliflower mosaic virus DNA: Possible involvement of RNA splicing and reverse transcription, EMBO J. 4, 1673-1680
- (52) Hogan, A. & Faust, E. A. (1986) Nonhomologous recombination in the Parvovirus chromosome: Role for a CT(A/T)TT(C/T) motif, Mol. Cell. Biol. 6, 3005-3009
- (53) Holl, J., Schmidt, C. & Schweyen, R. J. (1985) Cob intron 3 in yeast mtDNA: Nucleotide sequence and mutations in a novel RNA domain, in Achievements and Perspectives of Mitochondrial Research vol.II Biogenesis (Quagliariello E., Slater E.C., Palmieri F., Saccone C., Kroon A.M. eds.), Elsevier Science Publishers B.V. 1985, 227-236
- (54) Hu, N. & Messing, J. (1982) The making of strand-specific M13 probes, Gene 17, 617-626
- (55) Hudspeth, M. E. S., Ainley, W. M., Shumard, D. S., Butow, R. A. & Grossman, L. J. (1982) Location and structure of the *var1* gene on yeast mitochondrial DNA: Nucleotide sequence of the 40.0 allele, Cell 30, 617-626
- (56) Hudspeth, M. E. S., Vincent, R. D., Perlman, P. S., Shumard, D. S., Treisman, L. & Grossman, L. I. (1984) Expandable *var1* gene of yeast mitochondrial DNA: In-frame insertions can explain the strain-specific protein size polymorphisms, Proc. Natl. Acad. Sci. U.S.A. 81, 3148-3152

- (57) Inoue, T., Sullivan, F. X. & Cech, T. R. (1985) Intermolecular exon ligation of the rRNA precursor of *Tetrahymena*: Oligonucleotides can function as 5'exons, *Cell* 43, 431-437
- (58) Jacquier, A. & Dujon, B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene, *Cell* 41, 383-394
- (59) Jacquier, A. & Rosbash, M. (1986) Efficient trans-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5'exon-intron interaction, *Science* 234, 1099-1104
- (60) Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. (1984) Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose, *Gene Anal. Techn.* 1, 3-8
- (61) Johnson, R. C. & Melvin, S. I. (1985) Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer, *Cell* 41, 781-791
- (62) Kovokova, V., Irmelerova, J. & Kovak, L. (1968) Oxidative phosphorylation in yeast: IV. Combination of a nuclear mutation affecting oxidative phosphorylation with cytoplasmic mutation to respiratory deficiency, *Biochim. Biophys. Acta* 162, 157-163
- (63) Kruger, K., Grabowski, P. J., Zaug, A. J., Sanders, J., Gottschling, D. E. & Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclisation of the ribosomal RNA intervening sequence of *Tetrahymena*, *Cell* 31, 147
- (64) Lang, F. B., Burger, G., Thomas, D., Bandlow, W. & Kaudewitz, F. (1977) A simple method for large scale preparation of mitochondria from microorganisms, *Anal. Biochem.* 77, 110-121
- (65) Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steiz, J. A. (1980) Are sn RNPs involved in splicing ?, *Nature (London)* 283, 220-224
- (66) Li, M., Tzagoloff, A., Underbrink-Lyon, K. & Martin, N. C. (1982) Identification of the paromomycin-resistance mutation in the 15S RNA gene of yeast mitochondria, *J. Biol. Chem.* 257, 5921-5928
- (67) Macino, G. & Tzagoloff, A. (1980) Assembly of the mitochondrial membrane system: Sequence analysis of a yeast mitochondrial ATPase gene containing the *oli2* and *oli4* loci, *Cell* 20, 507-517

- (68) Macradie, I. G., Scott, R. M., Zinn, A. R. & Butow, R. A. (1985) Transposition of an intron in yeast mitochondria requires a protein encoded by that intron, *Cell* 41, 395-402
- (69) Mandel, H. I. G. A. (1970) Calcium dependent bacteriophage DNA infection, *J. Biol. Chem.* 243, 4543
- (70) Maniatis, T., Frisch, E. F. & Sambrook, J. (1982) *Molecular cloning. A laboratory manual.* Cold Spring Harbor Laboratory Press, New York
- (71) Martin, R. P., Bordonné, R. & Dirheimer, G. (1982) The paromomycin region in the yeast mitochondrial genome, in *Cell function and differentiation; part B*, New York 1982, 355-365
- (72) Maxam, A. & Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages, *Meth. Enzymol.* 65, 499-560
- (73) Maxam, A. L. & Gilbert, W. (1977) A new method for sequencing DNA, *Proc. Natl. Acad. Sci. U.S.A.* 74, 560-564
- (74) Messing, J. (1983) New M13 vectors for cloning, *Meth. Enzymol.* 101, 20-78
- (75) Messing, J., Crea, R. & Seeburg, P. H. (1981) A system for shotgun DNA sequencing, *Nucl. Acids Res.* 9, 309-321
- (76) Michel, F. & Dujon, B. (1983) Conservation of RNA secondary structures in two intron families including mitochondrial- chloroplast- and nuclear encoded members, *EMBO J.* 2, 33-38
- (77) Michel, F. & Lang, F. (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses, *Nature (London)* 316, 641-642
- (78) Michel, F. (1984) A maturase-like coding sequence downstream of the *oxi2* gene of yeast mitochondrial DNA is interrupted by two GC clusters and a putative end-of-messenger signal, *Curr. Genet.* 8, 307-317
- (79) Michel, F., Jacquier, A. & Dujon, B. (1982) Comparision of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure, *Biochimie* 64, 867-881

- (80) Miller, D. L., Underbrink-Lyon, K., Najarian, D. R., Krupp, J. & Martin, N. C. (1983) Transcription of yeast mitochondrial tRNA genes and processing of tRNA gene transcripts, in *Mitochondria 1983: Nucleo-mitochondrial interactions* (Schweyen R.J., Wolf K., Kaudewitz F. eds) , Walter de Gruyter Berlin - New York, 151-164
- (81) Morisato, D. & Kleckner, N. (1984) Transposase promotes double-strand breaks and single-strand joints at Tn10 termini in vitro, *Cell* 39, 181-190
- (82) Müller, M. (1985) Organisation, Prozessierung und Stabilität des tRNA^{glu}-COB Co-Transkriptes in Hefe Mitochondrien, Diplomarbeit , LMU München
- (83) Nargang, F. E., Bell, J. B., Stohl, J. B. & Lambowitz, A. M. (1984) The DNA sequence and genetic organisation of a *Neurospora* mitochondrial plasmid suggests a relationship to introns and mobile elements, *Cell* 38, 441-453
- (84) Nobrega, F. G. & Tzagoloff, A. (1980) Assembly of the mitochondrial membrane system: DNA sequence and organisation of the cytochrome b gene in *Saccharomyces cerevisiae* D273-10B, *J. Biol. Chem.* 255, 9828-9837
- (85) Novitski, C. E., Macreadie, J. G., Maxwell, R. J., Lukins, H. B., Linnane, A. W. & Nagley, P. (1984) Biogenesis of mitochondria: Genetic and molecular analysis of the *oli2* region of mitochondrial DNA in *Saccharomyces cerevisiae*, *Curr. Genet.* 8, 135-146
- (86) Ohtsubo e, E., Zenilman, M. & Ohtsubo, H. (1980) Plasmids containing insertion elements are potential transposons, *Proc. Natl. Acad. Sci. U.S.A.* 77, 750-754
- (87) Ohtsubo, E., Zenilman, M., Ohtsubo, H., McCorimick, M., Machida, C. & Machida, Y. (1981) Mechanism of insertion and cointegration mediated by IS1 and Tn3, *Cold Spring Harbor Symp. Quant. Biol.* 45, 283-295
- (88) Osiewacz, H. D. & Esser, K. (1984) The mitochondrial plasmid of *Podospora anserina*: A mobile intron of a mitochondrial gene, *Curr. Genet.* 8, 299-305
- (89) Osinga, K. A., Vries, E. D. E., Van der Horst, G. T. J. & Tabak, H. F. (1984) Initiation of transcription in yeast mitochondria: Analysis of origins of replication and of genes coding for a messenger RNA and a transfer RNA, *Nucl. Acids Res.* 12, 1889-1900
- (90) Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) Splicing of messenger RNA precursors, *Annu. Rev. Biochem.* 55, 1119

- (91) Pearston, D. H., Gordon, M. & Hardman, N. (1985) Transposon-like properties of the major, long repetitive sequence family in the genome of *Physarum polycephalum*, *EMBO J.* 4, 3557-3562
- (92) Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H. & Jarrell, K. A. (1986) A self-splicing RNA excises an intron lariat, *Cell* 44, 213-223
- (93) Perea, J. & Jacq, C. (1985) Role of the 5' hairpin structure in the splicing accuracy of the fourth intron of the yeast *cob-box* gene, *EMBO J.* 4, 3281-3288
- (94) Prunell, A. & Bernardi, G. (1974) The mitochondrial genome of wild-type yeast cells: IV. Genes and spacers, *J. Mol. Biol.* 86, 825-841
- (95) Prunell, A., Kopecka, H., Strauss, F. & Bernardi, G. (1977) The mitochondrial genome of wild-type yeast cells: V. genome evolution, *J. Mol. Biol.* 110, 17-52
- (96) Rigby, P. W., Dieckman, M., Rhodes, C. & Berg, P. (1977) Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I, *J. Mol. Biol.* 113, 237-251
- (97) Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467
- (98) Schweyen, R. J., Steyrer, U. & Kaudewitz, F. (1976) Mapping of mitochondrial genes in *Saccharomyces cerevisiae*, *Mol. Gen. Genet.* 146, 117-132
- (99) Schweyen, R. J., Weiss-Brummer, B., Backhaus, B. & Kaudewitz, F. (1977) The genetic map of mitochondrial genome including the fine structure of COB and OXI clusters., in *Mitochondria 1977: Genetics and biogenesis of mitochondria* (ed. Bandlow W. et al.), Walter de Gruyter, Berlin - New York
- (100) Schweyen, R. J., Weiss-Brummer, B., Backhaus, B. & Kaudewitz, F. (1978) The genetic map of the mitochondrial genome in yeast. Map position of drug^r and mit^- markers as revealed from population analysis of rho^- clones in *Saccharomyces cerevisiae*, *Mol. Gen. Genet.* 159, 151-160
- (101) Searles, L. L., Greenleaf, A. L., Kemp, W. E. & Voelker, R. A. (1986) Sites of P element insertion and structures of P element deletions in the 5' region of *Drosophila melanogaster* RpII215, *Mol. Cell. Biol.* 6, 3312-3319
- (102) Séraphin, B., Simon, M. & Faye, G. (1985) A mitochondrial reading frame which may code for a maturase-like protein in *Saccharomyces cerevisiae*, *Nucl. Acids Res.* 13, 3005-3014

- (103) Shapiro, J. A. (ed.) (1983) Mobile genetic elements, Academic Press: New York
- (104) Shapiro, J. A. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1933-1937
- (105) Simon, M., Zieg. J., Silverman, M., Mandel, G. & Doolittle, R. (1980) Genes whose mission is to jump. Phase variation: Evolution of a Controlling Element, Science 209, 1370-1374
- (106) Slonimski, P. P. & Tzagoloff, A. (1976) Localization in yeast mitochondrial DNA of mutations expressed in a deficiency of cytochromeoxidase and/or coenzyme QH2-cytochrome c reductase, Eur. J. Biochem. 61, 27-41
- (107) Sor, F. & Fukuhara, H. (1982) Nature of an inserted sequence in the mitochondrial gene coding for 15S ribosomal RNA of yeast, Nucl. Acids Res. 10, 1626-1633
- (108) Sor, F. & Fukuhara, H. (1982) Nucleotide sequence of the small ribosomal RNA gene from the mitochondria of *Saccharomyces cerevisiae*, in Mitochondrial Genes (Slonimski P. Borst P. Attardi G. eds), Cold Spring Harbor Laboratory Press, 255-267
- (109) Sor, F. & Fukuhara, H. (1983) Complete DNA sequence coding for the large ribosomal RNA of yeast mitochondria, Nucl. Acids Res. 11, 339-348
- (110) Sor, F. & Fukuhara, H. (1983) Unequal excision of complementary strands is involved in the generation of palindromic repetitions of rho- mitochondrial DNA in yeast, Cell 32, 381-396
- (111) Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98, 503-517
- (112) Stahl, U., Lemke, P. A., Tudzynski, P., Kück, U. & Esser, K. (1978) Mol. Gen. Genet. 162, 341-343
- (113) Steinhilber von, W. & Cummings, D. (1985) Reverse transcriptase activity in *Podospora anserina*, Abstract: International Bari Conference (Sept. 1985)
- (114) Stephen, C. W. & Kmer, A. (1985) Cleavage of cruciform DNA structures by an activity from *Sacharomyces cerevisiae*, Proc. Natl. Acad. Sci. U.S.A. 82, 6445-6449
- (115) Stümpke, H. & Steiner, G. (1981) Bau und Leben der Rhinogradentia, Gustav Fischer Verlag, Stuttgart, 19-20

- (116) Sylla, B. S., Huberdeau, D., Bourgaux-Ramoisy, D. & Bourgaux, P. (1984) Site-specific excision of integrated Polyoma DNA, *Cell* 37, 661-667
- (117) Tinoko Jr., I., Davis, R. C. & Jaskunas, J. R. (1968) in *Molecular Association in Biology*, Pullman, B. (ed), Academic Press, New York, 77
- (118) Tzagoloff, A., Nobrega, M., Akai, A. & Macino, G. (1980) Assembly of the mitochondrial membrane system: Organisation of yeast mitochondrial DNA in the *oli1* region, *Curr. Genet.* 2, 149-157
- (119) Uenatsu, Y., Kiefer, H., Schulze, R., Fischer-Lindal, K. & Steinmetz, M. (1986) Molecular characterization of a meiotic recombinational hotspot enhancing homologous equal crossing-over, *EMBO J.* 5, 2123-2129
- (120) van der Veen, R., Arnberg, A. C., van der Horst, G., Bohnen, L., Tabak, H. F. & Grivell, L. A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro, *Cell* 44, 225-234
- (121) Wakabayashi, K. & Mabuchi, T. (1984) Nucleotide sequence involved in the replication of cloned yeast mitochondrial DNA, *J. Biochem.* 96, 171-177
- (122) Waring, R. B. & Davies, R. W. (1984) Assessment of a model for intron RNA secondary structure relevant to RNA self-splicing: a review, *Gene* 28, 277-291
- (123) Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. (1986) Splice-site selection by a self-splicing RNA of *Tetrahymena*, *Nature (London)* 321, 133
- (124) Weiss-Brummer, B., Guba, R., Haid, A. & Schweyen, R. J. (1979) Fine structure of OX11: The mitochondrial gene coding for subunit II of yeast cytochrome oxidase, *Curr. Genet.* 1, 75-83
- (125) Young, I. G., Rogers, B. L., Campel, H. D., Jaworowski, A. & Shaw, D. C. (1981) Nucleotide sequence coding for respiratory NADH dehydrogenase of *Escherichia coli*. UUG initiation codon, *Eur. J. Biochem.* 116, 165-170
- (126) de Zamaroczy, M. & Bernardi, G. (1986) The GC clusters of the mitochondrial genome of yeast and their evolutionary origin, *Gene* 41, 1-22
- (127) de Zamaroczy, M., Faugeron-Fonty, G. & Bernardi, G. (1983) Excision sequences in the mitochondrial genome of yeast, *Gene* 21, 193-202
- (128) de Zamaroczy, M., Faugeron-Fonty, G., Baldacci, G., Goursot, R. & Bernardi, G. (1984) The ori sequences of the mitochondrial genome of a wild-type yeast strain: number; location; orientation and structure, *Gene* 32, 439-457

- (129) de Zamaroczy, M., Marotta, R., Faugeron-Fonty, G., Goursot, R., Mangin, M., Baldacci, G. & Bernardi, G. (1981) The origins of replication of the yeast mitochondrial genome and the phenomenon of suppressivity, *Nature* (London) 292, 75-78
- (130) de Zamaroczy, M. & Bernardi, G. (1986) The primary structure of the mitochondrial genome of *Saccharomyces cerevisiae* - a review, *Gene* 47, 155-177
- (131) de Zamaroczy, M. & Bernardi, G. (1985) Sequence organization of the mitochondrial genome of yeast -a review, *Gene* 37, 1-17
- (132) Zaug, A. J., Grabowski, P. J. & Cech, T. R. (1983) Autocatalytic cyclisation of an excised intervening sequence RNA is a cleavage-ligation reaction, *Nature* (London) 301, 578

An dieser Stelle möchte ich allen Mitarbeitern danken, die durch ihr Interesse und ihre Diskussionsbereitschaft zur Entstehung der vorliegenden Arbeit beigetragen haben.

Mein besonderer Dank gilt Herrn Prof. Dr. R. J. Schweyen für die Anregung der vorliegenden Untersuchung. Sein Rat und sein Interesse haben die Arbeit stets hilfreich begleitet.

Ferner bedanke ich mich bei:

Herrn Dr. R. Martin (Université Louis Pasteur), für die Mitteilung einer 2607 bp langen, an das 15S rRNA-Gen anschließenden Sequenz vor ihrer Publikation und für die Bereitstellung des Plasmids TR3-15A.

Herrn Dr. B. Dujon (Gif-sur-Yvette), für die Mitteilung der Sequenz eines im 15S rRNA-Gen optionalen GC-Cluster.

Frau C. Schüller und Herrn Dr. J. Zimmermann (Universität München), für fruchtbare Zusammenarbeit bei der Analyse optionaler GC-Cluster

Herrn H. Schmitz (Atelier für Werbung und Desktop Publishing, München) für die Bereitstellung eines Laserdruckers.

Dem Drachenflieger Club Aachental, für die Reproduktion meiner Schaffenskraft.

LEBENS LAUF

- 9.8.1953 Geburt in Weißenfeld (bei München), als erster von zwei Söhnen des Landwirts Georg Weiller und seiner Frau Karolina Weiller (geb. Gruber).
- Sept. 1959 Eintritt in die Volksschule Parsdorf.
- Juli 1967 Entlassung aus der 8. Klasse der Volksschule.
- Sept. 1967 Eintritt in die einjährige landwirtschaftliche Berufsfachschule in Pfarrkirchen (Internat) und gleichzeitiger Beginn einer landw. Lehre.
- Sept. 1968 Eintritt in die Berufsaufbauschule (BAS) in München Deroyst.1 (gewerblich technische Fachrichtung, Abendschule) und Fortsetzung der Lehre auf dem elterlichen Bauernhof.
- Sept. 1970 Erfolgreicher Abschluß der landw. Gehilfenprüfung.
- Juli 1971 Erfolgreicher Abschluß der Berufsaufbauschule (Mittlere Reife).
- Sept. 1971 Eintritt in die Berufsoberschule (BOS) München (naturwissenschaftlich technische Fachrichtung).
- Juli 1973 Erfolgreicher Abschluß der BOS und Erlangung der 'Fachgebundenen Hochschulreife'. Durch das Ablegen einer zusätzlichen Abiturprüfung im Fach 'Französisch' wurde damit die allgemeine Hochschulreife ermöglicht.
- Okt. 1973 Beginn eines Physik Studiums an der Technischen Universität München. Zugleich Beginn eines Abendkurses für Französisch.

- Sept.74 Abbruch des Physik Studiums und Übertritt an die Städtische Sprachenschule München (Herrenstr.).
- Juli 1976 Ablegung der Zusatzprüfung Französisch und Erlangung der allgemeinen Hochschulreife.
- Okt. 1976 Beginn eines Biologiestudiums an der Ludwig-Maximilians Universität München.
- Okt. 1983 Erfolgreicher Abschluß des Studiums mit dem Diplom (Fächerkombination: Genetik, Pflanzenphysiologie, Biochemie).
- Nov. 1983 Anstellung als wissenschaftlicher Mitarbeiter bei Prof. Dr. R. J. Schweyen.
- Juni 1987 Beabsichtigte Doktorprüfung im Fachbereich Biologie und Übertritt an die Universität in Dallas (USA).

